Hydrogen exchange mass spectrometry-based approaches for conformational analysis of disordered proteins in crowded conditions

By

Farai Ivan Rusinga

Submitted to the graduate degree program in Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson – David D. Weis, PhD

_____

Cindy Berrie, PhD

_____

Carey Johnson, PhD

_____

Michael Johnson, PhD

_____

Michael Wang, PhD

Date Defended: March 31, 2017

The Dissertation Committee for Farai Ivan Rusinga

certifies that this is the approved version of the following dissertation:

Hydrogen exchange mass spectrometry-based approaches for conformational analysis of disordered proteins in crowded conditions

_____

Chairperson David D. Weis, PhD

Date approved: April 18, 2017

Abstract

Disorder in proteins, or in extended protein regions, is more commonplace and functionally relevant than was conventionally assumed. Until a few decades ago, well-defined structure was considered as the only prerequisite for properly functional proteins. Since the development of disorder predictors, it has become apparent that intrinsically disordered proteins (IDPs) comprise significant proportions of eukaryotic genomes. An increasing number of investigations into the structure and function of IDPs reveal that they mediate a large number of crucial cellular processes and are associated with various disease processes. Despite the growing interest in IDPs, much remains unclear about the relationships between IDP structures, their interactions with other proteins and their mechanisms of action. Another important consideration is whether IDPs remain disordered in crowded conditions, such as the native cellular environments where IDPs function. Mass spectrometry (MS)-based approaches have been developed to probe IDP secondary structure and to complement other biophysical techniques in elucidating IDP functions. In addition, liquid-chromatography (LC) provides avenues for the removal of interference from complex protein samples containing high concentrations of polymeric crowding agents. Described herein, are studies based on hydrogen exchange (HX) measured by mass spectrometry (HX-MS) to: (1) propose a mechanism for calcineurin activation based on structural changes in its disordered regulatory domain, occurring upon binding calmodulin, (2) develop a method to probe the effects of polymer crowders on IDPs, and (3) investigate the effects of different crowder concentrations on transiently helical regions of random coil IDPs.

Calcineurin is a heterodimeric phosphatase, whose activity is regulated by calcium, calmodulin, and a regulatory subunit. The mechanism of calcineurin activation is directly

associated with binding of calmodulin to a disordered regulatory domain (RD) domain, located in the catalytic subunit. A calcineurin construct, comprising the RD, an autoinhibitory domain (AID), and a C terminal tail (referred to as the RD-AID-CT), was exposed to D_2O to compare HX in its free and calmodulin-bound forms. Free RD-AID-CT was fully exchanged at every residue after 5 s of exposure to D_2O. In the calmodulin-bound form HX was slowed in the calmodulin-binding domain and in an adjacent region. Slowed HX suggests that binding induces structure in the calmodulin-binding domain and an adjacent domain, in agreement with results from other biophysical techniques, such as CD and fluorescence anisotropy. The HX-MS data provided the spatial resolution required to localize calmodulin-induced structure to particular regions of RD-AID-CT. That localization provided enough information to propose a mechanism for calcineurin activation whereby the adoption of structure in the RD displaces the AID, exposing the active site cleft in the catalytic subunit.

Following the discussion of calcineurin's mechanism of activation, the development of a method for HX-MS analysis of CREB binding protein (CBP) in crowded buffers containing 300 g L^{-1} Ficoll, is outlined. CBP is a molten globular IDP consisting of three stable α-helices but a highly flexible tertiary structure. Ficoll is a synthetic polysaccharide used as an artificial crowder, but the high concentrations used to simulate cellular crowding present overwhelming interference to HX-MS analysis. To overcome this challenge, an LC method employing strong cation exchange (SCX) and reversed-phase extractions was developed to remove Ficoll before MS detection of CBP peptide segments. To validate the dual extraction approach, control experiments were performed using myoglobin in 300 g L^{-1} Ficoll samples, subjected to conventional HX-MS and SCX-based HX-MS workflows. Myoglobin peptide signals were lost in conventional HX-MS but were recovered after the SCX-based HX-MS workflows.

Unstructured myoglobin peptides were used to ensure that Ficoll neither alters the amounts of deuterium in HX labeling buffers nor affects the rates of deuterium uptake. Afterwards, HX-MS measurements showed that some CBP regions are stabilized while α-helices are destabilized in the presence of Ficoll.

The SCX-based HX-MS cleanup was applied to study crowding effects on a disordered domain of the activator of thyroid hormone and retinoid receptor (ACTR) in samples containing 300 g L^{-1} and 400 g L^{-1} of Ficoll. ACTR is a random coil IDP with transiently helical segments. Interestingly, all transiently helical regions of ACTR were destabilized in 300 g L^{-1} Ficoll, but one transiently helical region was stabilized in 400 g L^{-1} Ficoll. These results suggest that ACTR engages in non-specific interaction with Ficoll, disrupting intramolecular hydrogen bonds, in lower Ficoll concentrations. In contrast, ACTR is compacted by steric repulsions which can stabilize intramolecular hydrogen bonds in higher Ficoll concentrations. This information gives insights into the possible alterations of IDP structures, which could affect their functions, in different cellular compartments.

Acknowledgements

I would like to thank Dr. David Weis for generously sharing his knowledge, time and finances to make this work possible. Also, I am especially grateful for his availability, patience, understanding, encouragement, constructive criticism and humor, which fostered a comfortable and fulfilling work environment. I also thank the members of the Weis lab for being very helpful, inquisitive and fun group mates. Specifically, I would like to mention Dr. Theodore Keppel who kindly welcomed me and assisted with my orientation into graduate research. In addition, I thank Drs. Prakash Manikwar, Ranajoy Majumdar, Mohammed Al Naqshabandi, Jayant Aurora and Walid Maaty for scientific and other discussions that made the day-to-day easy. I also thank the current generation of Weis lab members, Tyler Hageman, Mihiri Weerasinghe, Chamalee Gamage, Juan Rincon and Yuqi Shi for valuable insights and for keeping me on my toes by seeking my advice. I would be remiss for failing to thank current members of the Volkin group as well, for their valuable assistance.

I am grateful to my dissertation committee for agreeing to evaluate my progress and to provide fair critiques that will undoubtedly nourish my advancement as a scientist. In addition to advice on the dissertation, I have received classroom and/or comprehensive orals instructions from all committee members. I will cherish all pieces of information and advice obtained through their instruction, which will prove invaluable in my career as an Analytical Chemist. I will also fondly remember Dr. Craig Lunte for the energy in educating me during all interactions I had with him. There are numerous other KU faculty members from the Chemistry Department, especially, who have always shown kindness and willingness to assist me in my graduate education. Drs. Todd Williams and Justin Douglas assisted immensely towards my projects and always shared a laugh in the hallways. I cannot forget to mention the administration, in

particular, the International Students Services for ensuring that other requirements outside of Chemistry were met, without which I could not successfully complete my studies.

Outside of KU, I would like to extend gratitude to Dr. Elaine Marzluff, my undergraduate advisor, for introducing me to protein structural analysis using hydrogen exchange mass spectrometry. She has maintained interest in my progress, continued to encourage me and continues to offer assistance towards my personal and professional well-being. I thank Dr. Trevor Creamer who worked with us on the calcineurin project and graciously offered his blessing for the reproduction of work in Chapter 2 of this dissertation. I would like to thank Peter Smith from LEAP Technologies for assistance with programming the automated SCX workflow, and continued technical support throughout my projects. My interest in science was sparked by the enthusiasm and personal encouragement from my Physics, Chemisty, Biology and Mathematics teachers at my high school, St. Ignatius College in Zimbabwe. I will forever be indebted to their bending me in the direction of logic and *never assuming* facts.

Of course science does not operate in a vacuum. Therefore, I must thank all other instructors along my life and academic journey for shaping my sensitivity to various other subjects. Special mention goes to my dear partner Nichole Carmen who has sailed the waters, borne the brunt and always aided in steering the ship, which is the life of a graduate student. Together we have raised our beautiful daughter, Auralai Carmen Rusinga, who has brought much joy (and gentle frustrations) for the past three years. They will both no doubt continue to motivate me in my career.I thank my dear friend Jacob Newell for life lessons, encouragement and therapeutic woodworking lessons. I thank my other friends who have always showed great belief in me even when self-doubt set in.

I dedicate this work to my family who raised me, saw a hunger for knowledge and fostered my efforts towards satisfying that hunger. I thank my late father, Andrew Rusinga, whose intelligence I can only try to emulate, for teaching me to seek knowledge in books and to learn many skills from my earliest days. I thank my mother, Grace Tsamba, for her unquestionable belief in me. I thank my adoptive parents, David and Svetlana Rusinga, who worked tirelessly to provide all financial and other support for me to study abroad. I thank all uncles and aunts who saw a promising young man and never hesitated to share that thought and encouragements. In my life and academic journey I have encountered many *angels from God* and I dedicate this work to all of them!

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Intrinsically disordered proteins

1.1.1 Disorder versus the structure-function paradigm

The 1950s ushered in the era of modern molecular biology, with the solving of both the DNA double-helix and the first atomic structures of proteins, myoglobin[1] and hemoglobin[2], using X-ray diffraction patterns. In 1962, Watson and Crick were awarded the Nobel Prize for Physiology and Medicine for work on the DNA structure. That same year, John Kendrew and Max Perutz shared the Nobel Prize for Chemistry for the efforts on protein structures. Further progress was made by Christian Anfinsen, who showed that denatured ribonuclease spontaneously refolds and regains function when replaced into non-denaturing conditions, suggesting that protein three-dimensional structure is encoded in the amino acid sequence.[3] For this work, he too earned the Nobel Prize for Chemistry, in 1972. These advances supported the prevailing *central dogma* of structural biology, which assumes that well-defined structure is essential for protein function. However, intrinsically disordered proteins (IDPs) have since been recognized as a class of proteins which, instead, derive function from being unstructured.[4-9] Unlike structured proteins, IDPs comprise extended regions or entire proteins lacking stable secondary or tertiary structure.[7, 10]

1.1.2 The protein order-disorder continuum

Based on the extent of disorder, proteins can be classified into broad categories along a protein order-disorder continuum. These categories form the so-called quartet of protein states, namely ordered, molten-globule, pre-molten globule and intrinsic coil proteins.[11] Ordered proteins fold into stable secondary and tertiary structures. Their native-like ensemble states deviate minimally from well-defined conformations.[12] Molten globular proteins are a broad and

nebulous class, with stable secondary structure but highly flexible tertiary structure.[13-15] Pre-molten globules have *fairly* stable secondary structure, which is more frequently disrupted than the secondary structure of molten globules or ordered proteins.[16-18] Intrinsic coil proteins are predominantly unstructured under native conditions. Some intrinsic coils can undergo transient structure formation, which persists within their unstructured ensembles and is sometimes stabilized when bound to protein ligands.[14, 15, 17, 19] To further expound, the order-disorder continuum is a spectrum and functional proteins exist all along this spectrum.

1.1.3 Disorder predictions and prevalence of protein disorder

Numerous disorder-prediction algorithms are used to predict the prevalence of IDPs in different genomes, even before specific IDPs are characterized experimentally.[20] IDPs exhibit low sequence complexity, with their primary structures composed mainly of charged[21-23] and polar residues. This sequence composition is predicted to promote a high degree of peptide chain flexibility.[24-26] Disorder predictors are based on protein sequence, but other residue characteristics e.g. hydropathy and bulkiness have enhanced the accuracy of disorder prediction algorithms.[10, 27, 28]Two examples of disorder predictors are PONDR[25, 29], which was trained using databases of structured and disordered proteins and FoldIndex, which is based on amino acid net charge and average hydrophobicity.[30] With expanding protein databases and ever-increasing computing power, predictive abilities continue to improve with every iteration of algorithms, such as PONDR.[10, 31, 32]

The application of disorder predictors to genome-wide databases suggests that IDPs occur more frequently across the animal kingdom than was earlier supposed.[31] These algorithms have revealed that about 44% of proteins coded by the human genome possess long (> 30

residues) loops, disordered end-terminals, hinge regions, domains or entire sequences that are disordered.[31] Consensus predictions, which combined and compared predictions from different disorder predictors, suggest that IDP frequencies range from around 5% of *E. coli* proteins to 28% of *M. musculus* proteins of mostly disordered proteins.[28] Apparently, protein disorder is less common in prokaryotes, with only 2% and 4% putative disordered regions for proteins from archaean and eubacterial genomes, respectively.[33] With a structure-oriented view of protein function, the prevalence of IDPs was initially surprising. However, evidence suggests IDP functions are vital, especially for multicellular organisms.

1.1.4 IDP functions

Ironically, efforts to predict and classify protein functions based on well-defined structure (as a prerequisite for function) led to the discovery that IDPs are very prevalent in eukaryotic genomes.[31] Far from being passive proteins or aberrant domains, it is now clear that many IDP functions are facilitated by their lack of well-defined native structure.[34] To reiterate, structured proteins are also flexible but the native-like conformers are all very similar. The, static three-dimensional illustration, obtained from X-ray crystal structures and customarily presented in textbooks, can be somewhat misleading.[35]The main difference is that IDP ensembles consist of diverse, unstructured, conformations compared to ordered proteins. The unfolded structures of IDPs are an extension Anfinsen's conclusions because disorder (as much as spontaneous folding) is also encoded in IDP sequences. Unlike ordered proteins, such as ribonuclease, IDPs are depleted in large hydrophobic residues (W, F, I, Y, V, L and M) around which hydrophobic cores would form.[4, 7Bae, 2009 #349, 24, 29, 36] Therefore, IDPs and structured proteins evolved their functional properties differently, with each protein class being *purpose-built* for its respective functions.[34, 37]

IDP functions, that are facilitated by disorder include (but are not limited to) regulation via post-translational modifications, recruitment of different binding partners, and conformational variability and adaptability.[38, 39] Some IDPs, such as entropic chains, benefit directly from disorder and remain flexible under all native conditions.[10, 40]In contrast, effectors sometimes adopt stable structure upon binding, consequently switching between inactive and active forms or activating their binding partners.[14] In Chapter 2 of this dissertation, a mechanism is proposed for activation of calcineurin, involving structuring of its disordered regulatory domain upon binding calmodulin. Structural plasticity allows some IDPs to bind multiple binding partners in assemblies[34]. Functional plasticity allows IDPs to bind different binding partners in the same binding site, adopting different folds to elicit different effects.[9] More intriguing yet, some IDPs utilize disorder in different ways, e.g. CREB, whose disordered domains simultaneously act as effectors and assemblers of transcription factors in the nucleus.[14, 34, 41, 42]. Given myriad of cellular functions mediated by IDPs, it is hardly surprising that their overexpression and misfolding is implicated in numerous disease conditions.

1.1.5 Roles of IDPs in diseases

The unique structural features, physical properties and regulatory functions of IDPs necessitates their tight regulation in cells because dysregulation or misfolding can have adverse consequences.[39, 43, 44] Many disease conditions have been classified as *diseases of disorder* because of their link to IDPs.[45]For example, the IDPs that are involved in chromosomal translocation are implicated in several types of juvenile leukemia and other cancers.[18, 44, 46] The propensity for IDPs to aggregate is directly associated with neurodegenerative conditions like Alzheimer's diseases and Parkinson's disease.[47, 48] In some cases, a single IDP has been associated with several unrelated diseases. For example, overexpression of calcineurin, an

enzyme which is broadly distributed in tissue, is linked with cardiac hypertrophy[49]. Conversely, deficiency of calcineurin, resulting from the use of of calcineurin inhibitors after organ transplants, results in the onset of diabetes mellitus[50]. On the other hand, mice expressing mutated, inactive, calcineurin exhibited impaired memory and symptoms consistent with schizophrenia.[51] With increasing evidence of IDP involvement in diseases, unstructured protein domains have become rational targets for small molecule drugs.[52-55] Perhaps IDPs could also become drug candidates, building on work involving peptide short linear motifs or SLiMs.[56] IDPs are clearly a medically-relevant protein class and understanding their sensitivity to their local cellular environments, given their structural plasticity, could be key in elucidating the causes of dysregulation or misfolding.

1.1.6 Calcineurin

Chapter 2 of the dissertation describes a study on a disordered region of calcineurin (CaN). CaN is a serine/threonine phosphatase that is activated by calcium and calmodulin. CaN is heterodimeric, comprised of CaN A, a calmodulin-binding catalytic domain, and CaN B, a

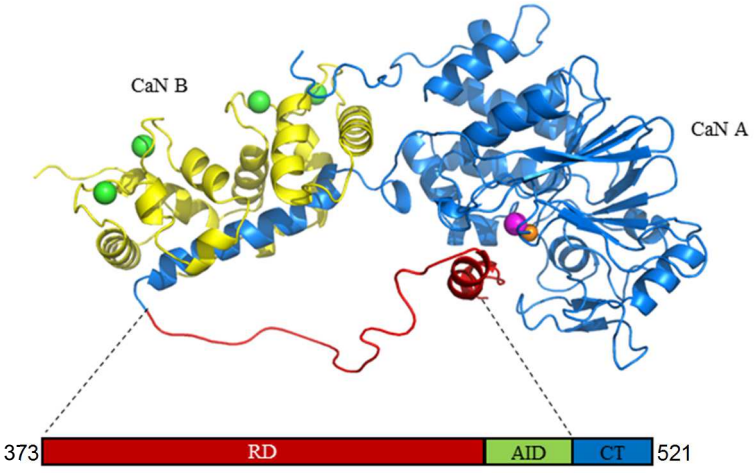


Figure 1.1 Full length calcineurin showing catalytic CaN A and regulatory CaN B and auto-inhibitory domain (PDB entry 1AUI[57]). The disordered RD and CT are missing from the X-ray crystal structure. The random coil representing the RD (red) was manually added to PDB entry 1AUI. The domain organization of the RD-AID-CT construct is shown.

calcium-binding regulatory domain. The calmodulin-binding, regulatory domain (RD), of CaN A is disordered (see Figure 1.1). Hydrogen exchange mass spectrometry, along with other biophysical techniques, was used to study a construct of human CaN. The construct consists of the RD, an autoinhibitory domain (AID), and a C-terminal tail (CT), representing residues 373-521 of CaN A (hereafter referred to as RD-AID-CT, see Figure 1.1 based on calcineurin X-ray crystal structure[57]). HX data, from chapter 2 of this dissertation, shows that segments of RD-AID-CT form stable secondary structure upon binding calmodulin. Along with other biophysical techniques, the HX-MS results were key in proposing a mechanism for activation of CaN.

1.2 Macromolecular crowding

1.2.1 Volume exclusion and non-specific interactions in crowded conditions

Biological fluids, such as cytoplasm, contain large quantities of macromolecular co-solutes. For example, the interiors of eukaryotic cells contain about 300 g L^{-1} of macromolecules such as proteins, polypeptides, nucleotides, polysaccharides, and extended networks of macromolecular structures and membranes.[58-61]. Steric repulsions and non-specific interactions between proteins, and other macromolecules, forced into close proximity by the limited intracellular space, are called *macromolecular crowding effect*s.[62-65]. Although macromolecular crowding could influence protein structure and dynamics, altering protein function, the vast majority of protein conformational studies are performed in dilute homogeneous buffers.

In the early 1960s, Ogston and Laurent reported that solutions containing high concentrations of macromolecules (hyaluronic acid and dextran) markedly affect the diffusion rate of proteins, due to volume exclusion.[66-68] Furthermore, they observed that these effects become more pronounced as the size and concentration of the *crowders* increases. To clarify an

important point, the macromolecular crowders, referred to here, are *background* species which do not participate in direct reactions or specific interactions with the proteins.[61] In the 1980s, Minton first showed that macromolecular crowding affects protein conformations and stability, via steric repulsions (also known as volume exclusion), which in turn affects protein function.[62, 69, 70] Since then, several researchers have shown that crowding could decrease the rate of signal transduction and metabolite channeling[71], perturb the structure and function of enzymes[72], improve efficiency of chaperone mediated protein folding[73] and promote protein association[74]. Some of the processes listed here are associated with IDP functions, yet IDP structure and function are typically studies in dilute solutions.

Initially, macromolecular crowding effects on protein conformations were studied by adding polymers, such as dextran, only to exclude volume in protein solutions. Crowding effects are now also studied *in silico* and *in vivo*.[61, 63, 64, 75-78] Most simulations of crowded conditions are based on volume exclusion theory, which treats macromolecular crowders as inert, hard spheres. Indeed, volume exclusion is the predominant effect observed for structured proteins in crowded conditions.[63, 79]Volume exclusion (also called hard interactions) occurs because proteins cannot occupy the volume that is occupied by crowding agents. Therefore, compact conformations of ordered proteins, which take up less space, are entropically favored.[63, 80] However, crowders are not truly inert. As a result, non-specific interaction (or soft interactions) can occur between proteins and crowders.[81] The crowding macromolecules used for *in vitro* studies include proteins (e.g. albumin[82]), cell lysates[65], and synthetic polymers, such as Ficoll (a polysaccharide), dextran, polyethylene glycol (PEG) and polyvinylpyrrolidone (PVP).[72, 81, 83, 84].

Soft interactions depend on the protein-crowder pair and include charge interactions, dipole-dipole interactions, hydrophobic interactions and hydrogen bonding. Soft interactions can

be stabilizing or destabilizing, depending on the physico-chemical properties of the protein and the crowder. For example, the globular protein CI2 is compacted differently by albumin compared to Ficoll.[82] Moreover, globular proteins are always compacted by polymer crowders[81] because their native-like conformations fold around hydrophobic cores, avoiding crowders. In contrast, IDPs could be destabilized, under similar crowding conditions, because they prefer extended conformations, making them more likely to engage in soft interactions.[84] Furthermore, polymer crowders are not self-avoiding hard spheres. Instead, their chains interpenetrate and form mesh-like networks above certain concentrations[85], increasing their likelihood of interacting with IDPs.

In-cell crowding studies reveal that proteins can also engage in so-called *quinary* interactions with the intracellular matrix, which could modulate protein stability.[86]These observations are supported by results from *in vitro* experiments, which showed that charge interactions between proteins and proteins in *E. coli* lysates affect protein stability, depending on the protein's isoelectric point.[65] Although observations depend on the protein-crowder pair, both hard and soft interactions are always present in crowded conditions. Furthermore, the structural plasticity of IDPs complicates effort to model macromolecular crowding effects on IDPs, particularly in the complex intracellular environment.

1.2.2 Potential influence of intracellular crowders on IDP structure and function

Insufficient theoretical frameworks necessitate experimental approaches towards understanding interactions between IDPs and cellular macromolecules. Such interactions could modulate IDP conformations, potentially altering their functions. Unlike globular proteins, whose native-like conformations are generally stabilized by volume exclusion, IDPs could be

more susceptible to changes in crowded environments. For example, disordered transcription factors, such as ACTR, could adopt conformations that promote binding to chaperones for transport into the nucleus under the influence of cytosolic (mostly globular) proteins. In the nucleus, ACTR would experience different crowders, i.e. negatively-charged and extended nucleic acids, which could alter ACTR conformational ensembles to promote release from the chaperone and binding to other transcription factors and DNA. Such crowder-regulated binding could minimize non-productive binding, increasing the efficiency of ACTR transfer and activation. Crowding could also affect rates of post-translational modification, control assembly of protein complexes, regulate enzymatic digestion, or change the capture radius for IDP targets. An interesting case is that of prothymosin-α (ProTα), a highly-charged intrinsic coil in dilute solution. ProTα is dramatically compacted in in 40 % PEG solutions, where its radius of gyration decreases by 30 % (compared to about 10% more compact IDPs).[84] This contrasting compaction behavior, for different IDPs, calls into question the accuracy of inferences drawn from dilute solution studies about the behaviors of some IDPs' inside cells.

Investigations of macromolecular crowding effects on IDP folding could aid in elucidating the mechanisms of dysregulations and misfolding, towards understanding the link between IDPs and disorder diseases. Diseases linked to IDPs could arise from changes in cytosolic compositions resulting from factors like malnutrition or other seemingly unrelated conditions. For example, the aggregation of IDPs (e.g. α-synuclein in Lewry bodies found inside the neurons of patients with Parkinson's disease[87]) could be linked to changes in the macromolecular composition of the intracellular matrix. Such changes could result from leakage of cellular components, due to cell damage, or from the introduction of foreign species, which may have crossed the blood-brain barrier. Modeling intracellular crowding effects, especially in

dynamic cellular environments, is complicated by the diversity and plasticity of IDP ensembles and the complex interplay between hard and soft interactions in crowded environments. Observations from parallel *in vitro* studies, employing different concentrations or combinations of synthetic polymer crowders, can aid in drawing inferences about crowding effects on IDP conformations in different cellular compartments. Such inferences could assist in distinguishing primary causes from coincidental effects of disorder diseases, aiding in designing suitable interventions against those diseases.

1.2.3 IDP models: ACTR and CBP

Chapters 3 and 4 of this dissertation describe Ficoll-induced crowding effects on two model IDPs, on different ends of the order-disorder continuum. Chapter 3 outlines crowding studies on the nuclear coactivator binding domain (NCBD) of mouse cAMP-responsive element (CRE)-binding (CREB) protein (residues 2059-2117), hereafter referred abbreviated CBP. CBP is a molten globular IDP, with three stable α-helices but an ill-defined tertiary structure (see Figure 1a).[14, 15, 88, 89]. CBP was used as a model system to evaluate a strong cation exchange (SCX)-based method to remove crowding agents for hydrogen exchange mass spectrometry (HX-MS) analysis of proteins in crowded conditions. Chapter 4 outlines crowding studies on a domains of human p160 coactivator, activator of thyroid and retinoid receptor (residues 1023-1093, hereafter abbreviated ACTR. ACTR is an intrinsic coil IDP, with region that transiently form α-helices (see Figure 1b).[14, 15, 90, 91] ACTR was previously observed to become compact in concentrated PEG solutions, but with unexpected PEG concentration and molecular weight dependence.[84] In chapter 4, Ficoll effects on ACTR transiently helical regions are investigated in two different Ficoll concentrations.

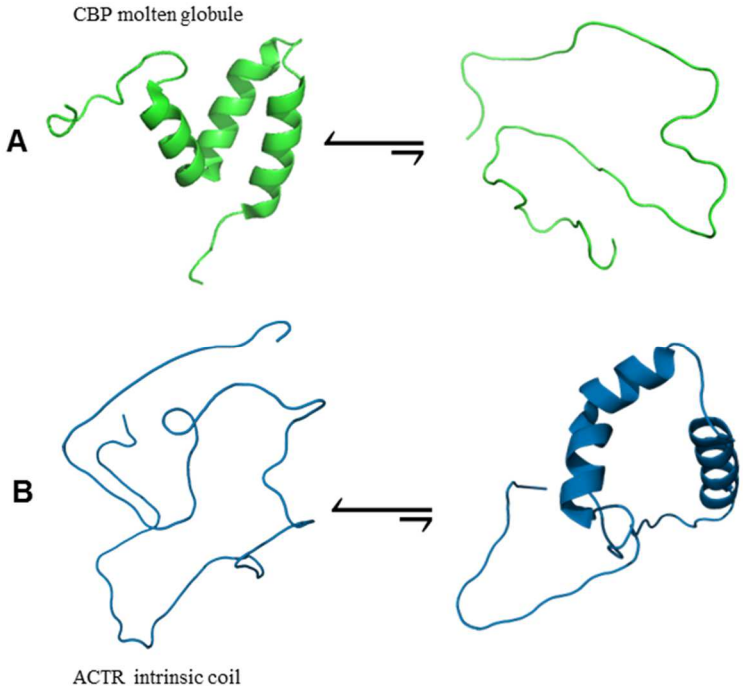


Figure 1.2 (A) CBP is a molten globule with stable α-helices that transiently unfolds and (B) ACTR is an intrinsic coil that forms transient α-helices. The representations of CBP and ACTR are based on PDB entry 1KBH[15]

1.3 Methods for characterizing IDPs

1.3.1 Identifying disorder in proteins

Some general properties of IDPs, such as lacking hydrophobic cores, distinguish them from ordered proteins. Biophysical techniques which measure quantities related to these properties can aid in identifying disordered protein domains. These techniques include extrinsic florescence spectroscopy (EFS), limited proteolysis, differential scanning calorimetry (DSC) and immunochemistry. EFS measures extrinsic florescence of compounds (e.g. 1-anilino-8-naphthalenesulfonic acid), which fluoresces when bound to non-polar surfaces such as residual hydrophobic cores of molten globular or pre-molten globular IDPs.[92] Limited proteolysis exploits the increased susceptibility towards enzymatic digestion, of flexible regions, to distinguish ordered and disordered proteins. For example, pre-molten globular apomyoglobin is

digested under limited exposure to trypsin but globular holomyoglobin is resistant to digestion under similar.[93] DSC measures differences in the amount of heat required to raise the temperature of a protein sample compared to a reference, as a function of temperature.[94]The absence of heat absorption peaks in DSC curves indicates protein disorder because IDPs do not undergo thermal unfolding.[95] Immunochemical approaches utilize antibodies that recognize conformational epitopes i.e. epitopes consisting of residues that are proximal in tertiary structure but distal in secondary structure.[94] Flexibility in the tertiary structure of an IDP antigen could separate residues, otherwise comprising its conformational epitope, thereby diminishing antibody binding.

Techniques which measure overall protein dimensions can also aid in identifying disorder, because IDPs tend to have extended conformations compared to ordered proteins of equivalent mass. Example of these techniques include, small angle X-ray scattering (SAXS), fluorescence correlation spectroscopy (FCS), dynamic light scattering (DLS), and size-exclusion chromatography (SEC). SAXS is a solution-based method that measures the angles of X-rays scattered by electrons to determine average radii of gyration of IDP ensembles. In contrast to compact proteins, extended proteins have reduced electron densities resulting in distinct SAXS (Kratky) plots.[96, 97] FCS measures the fluorescence of molecules in highly diluted solutions as they randomly pass through a narrow optical window illuminated by an excitation laser. Fluorescence fluctuates due to Brownian motion, which can be correlated with diffusion rates of fluorophore-tagged IDPs to compute their hydrodynamic radii.[98] DLS measures the interference pattern of light scattered by molecules undergoing Brownian motion. The interference pattern correlates with translational diffusion coefficients, from which the Stokes' radii of IDPs can be determined using the Stokes-Einstein equation.[99, 100] SEC measures the mobility of proteins

through a gel matrix in a column based on the proteins' hydrodynamic radius. Larger proteins elute in lower elution volumes because they are retained less than smaller molecules in the gel matrix. Owing to their extended conformations, IDPs have higher *apparent mass* than globular proteins of similar molecular weight. Using pre-determined molecular mass-to-radius of gyration ratios, SEC can provide information about the compactness of a protein.[101] The methods described above rely on anticipating physico-chemical properties indicative of protein disorder but they do not provide secondary or tertiary structure information.

1.3.2 Secondary and tertiary structure determination

Transient secondary and tertiary IDP structures are challenging to characterize because many biophysical techniques specifically detect stable secondary and tertiary structures. In addition, high concentrations of macromolecular crowders, orders of magnitude above IDP concentration, introduce overwhelming interference that can masks IDP signals. However, sensitive techniques like circular dichroism (CD), nuclear magnetic resonance (NMR), Förster resonance energy transfer spectroscopy (FRET) and mass spectrometry (MS) have made IDP conformational analysis more accessible, even in crowded conditions. CD, FRET, and NMR will now be discussed briefly, in the context of IDP conformational analysis and macromolecular crowding studies. MS, the key technique employed in the studies outlined in the remaining chapters, will be discussed more extensively thereafter.

1.3.2.1 Circular dichroism spectroscopy

CD spectroscopy measures the differential absorption of right- and left-handed circularly polarized light by chiral molecules, such as the amino acids in proteins.[102] In the far-UV range (190-230 nm) CD is used to determines the proportions of secondary structure elements in a

protein, because amino acids in different local environments absorb different wavelengths of circularly polarized light. Specifically, α-helices exhibit negative absorption bands at 222 nm and 208 nm, and a positive band at 193 nm; β-sheets have a negative band at 218 nm and a positive band at 195 nm; and intrinsic coils have low absorption above 210 nm and a negative band near 195 nm.[103-105] CD measurements are relatively simple and quick, and are commonly used to confirm protein folding after expression and purification. CD revealed that both tau protein[106] and α-synuclein[107](which form insoluble aggregate in neurons of patients with Alzheimer's disease) lack significant amounts of secondary structure. CD data shows that the α-helical content of CBP and ACTR, the model IDPs discussed in chapters 3 and 4, increases when they bind to each other.[14] Koutsioubas and others used CD to show that helicity is disrupted by the polymer crowder PEG.[108]CD was also used to monitor changes in helical content of the disordered domain of c-Fos under crowded conditions.[109] While CD can be used to quantify the fraction of residues in different secondary structural elements, it does not provide localized structural information along a protein sequence. Also, CD cannot distinguish between individual proteins in a mixture or between conformers in an ensemble. Instead CD averages the secondary structure elements of all components of a protein mixture or ensemble.

1.3.2.2 Förster resonance energy transfer spectroscopy

FRET spectroscopy measures non-radiative energy transfer which occurs via long-range dipole-dipole coupling, from a fluorophore in an excited electronic state, called the *donor* (D), to another chromophore, called the *acceptor* (A).[110, 111] FRET efficiency is inversely proportional to the sixth power of the distance between the D-A pair, so FRET efficiency can be correlated with protein (global) conformational changes between 1-10 nm. Single molecule FRET revealed that IDPs become compact with increasing concentrations of polymer crowders.[84] FRET

measurements can be correlated with known protein structures, using molecular modelling, to determine changes in protein shape.[72] A limitation of FRET is that it does not probe secondary structural changes in a single measurements, but such information can be inferred if multiple FRET pairs are tethered to different protein segments.[112] In addition, bulky chromophores could influence IDP dynamics in crowded conditions. Also, polymer crowders must be of sufficient purity to avoid interference from contaminants which could absorb excitation light and/or produce background florescence at physiologically-relevant concentrations.[84]

1.3.2.3 NMR spectroscopy

NMR spectroscopy measures the magnetic field properties of atomic nuclei possessing spin $\frac{1}{2}$ nuclei, such as ^{1}H, ^{13}C and ^{15}N. NMR is capable of atomic-scale resolution, providing information about the local chemical environment and spatial arrangement of each NMR active nucleus within a protein.[113, 114]Therefore, NMR provides residue-level secondary and tertiary structure information, unlike FRET and CD.[115] In fact, NMR is arguably the most useful technique for probing IDP structural changes, in both dilute and crowded conditions.[78]Transiently helical regions of ACTR have been identified using NMR secondary shifts, a comparison of chemical shifts for IDPs under native versus denaturing conditions.[116] NMR measurements showed that α-synuclein remains disordered in crowded *E. coli* periplasm.[77] It has been determined that charge interactions can alter the structure of proteins crowded by cell lysates (which were pre-labeled with non NMR-active deuterium to prevent interference).[65] Conversely, hydrogen-deuterium exchange of ^{15}N labeled protein, in crowded D_2O buffers, was measured using heteronuclear single quantum correlation (HSQC)-NMR to detect crowder-induced conformational changes.[117] Despite its undoubted merits, NMR is generally limited to proteins with masses below 35 kDa because larger proteins tend to have overlapping resonances.

Also, NMR typically requires milligrams of protein, quantities which are not always readily available. In addition, IDP resonances can be lost in crowded solutions because the rotational diffusion of extended molecules is significantly slowed, which increases rotational correlation times and broadens NMR peaks.[118]

1.4 IDP conformational analysis using mass spectrometry

1.4.1 Protein mass spectrometry

Mass spectrometry (MS) measures the mass of charged molecules, with high precision[119]. Initially, MS measurements were unsuitable for protein structural analyses because previously-available ionization methods would certainly fragment proteins. The invention of *soft ionization* techniques, such as electrospray ionization (ESI)[120] and matrix-assisted laser desorption ionization (MALDI)[121], has facilitated MS detection of intact proteins and peptides. MS now complements and (in many cases) provides an efficient alternative to NMR spectroscopy and X-ray crystallography for medium-to-high resolution protein conformational analyses.

1.4.2 Native electrospray ionization

Native ESI-MS takes advantage of soft ionization to preserve protein structure during the transition from solvent to gas phase, which occurs at the ESI source. For native ESI-MS, volatile buffer salts (e.g. ammonium acetate) are used to maintain neutral pH, and to aid in desolvation and ionization.[122] Native ESI is particularly useful for MS analysis of IDPs because extended conformations typically present more solvent-exposed ionization sites resulting in distinct charge-states compared to compact proteins[123]. Charge-state distribution can reveal the presence and the respective proportion of conformers within protein ensembles.[124, 125] For example, the

pH-dependence of α-synuclein conformers was determined using charge-state distribution in MS spectra.[126]

The capabilities of native ESI, to resolve conformers, can be enhanced by incorporating rapid, gas phase, ion mobility (IM) separation. IM-MS measures the mobility of charged molecules in a drift tube, where ions are accelerated by an electric field and experience drag from an inert drift gas.[119] Therefore, IM separates conformers based on their collision cross section, which is related to their shape and size. An advanced IM-MS approach called field asymmetric ion mobility spectrometry (FAIMS) provides unprecedented resolution of conformers, possessing only the slightest structural differences.[127] A limitation of native ESI is that it can only identifies the presence of conformers, without resolving specific structural features. Furthermore, electrostatic repulsions could force highly-charged proteins to unfold and adopt extended conformations. As a result, the preservation of protein structure in the gas phase is a contentious assumption.[128] Finally, native ESI of proteins in crowded solutions is impractical with current workflows because of potentially overwhelming interference and instrument contamination.

1.4.3 Solution-based labeling methods for mass spectrometry

To remove the constraint of solution-phase structure preservation during ESI and to facilitate conformational analysis in complex matrixes[129-131], solution-based labeling can be performed before ESI-MS. Solution-based labeling is the application of covalently-bound probes to proteins. The incorporated labels change proteins mass by a predictable amount, which can then be measured using MS. Solvent-exposed or flexible protein segments undergo labeling reactions at labile sites faster than buried or rigid segments. Therefore, labeling reveals structural

information and time-resolved dynamical information. Two common labeling techniques are hydroxyl radical labeling[132-134] and chemical cross-linking[135, 136]. These techniques have been extensively reviewed elsewhere. Here, other amino acid labeling methods, which have been used to study disordered proteins, will be discussed.

Amino acids e.g. arginine (R), carboxylic acids (D and E), cysteine (C), histidine (H), lysine (K), tryptophan (W) and tyrosine (Y) can be modified by reactions with derivatization reagents.[137]From that list, R, E and K are particularly abundant in IDPs.[36] R reacts with vicinal dicarbonyl compounds, e.g. 2,3-butanedione (adding 86 Da per modification), to form cyclic adducts, in neutral pH and at room temperature. R labeling is suitable for native structural analysis; carefully-timed R labeling has been used to determine the number of exposed R sites in apomyoglobin.[138] Different R labeling reactions can correlate differently (either directly or inversely) with solvent accessibility, providing a means to identify ordered and disordered regions of the same protein.[137] Carboxylic acids can be labeled with carbodiimides e.g. 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC), below pH 6 and in the presence of a nucleophile.[139] Selective carboxylate labeling was used to show that the effector region of active Ras-GTP protein is more disordered than inactive Ras-GDP.[140] Acetylation of K is achieved using organic anhydrides, or *N*-hydroxysuccinimide derivatives, in neutral to basic pH. At neutral pH, almost all K residues are positively charged, and K is associated with the binding sites of transcription factors. Transcription factors are typically IDPs, which bind to negatively-charged DNA.[137] The binding surface of a thyroid transcription factor–DNA complex was characterized using a combination of limited proteolysis and selective acetylation detected by MS.[141]

After labeling, intact proteins can be delivered directly to ESI-MS to provide global structural information especially for labels with high amino acid specificity. For example, 2,3-butanedione and phenylboronic acid modification of arginine was used to show differences in the compactness holomyoglobin and apomyoglobin.[138] Spatial resolution can be achieved by fragmenting the protein using enzymatic digestion before ESI, in what is termed *bottom-up* MS. Fragmentation can also be performed after ESI using electron transfer (ET) or electron capture (EC) dissociation, an approach called *top-down* MS, or by using a combination bottom-up and top-down MS, called *middle-down* MS.[142, 143] Peptide-level resolution is achieved in bottom-up MS, but can be enhanced in top- and middle-down MS or by combining results from proteolysis by different enzymes.[144] Fragmentation localizes labeled sites, allowing for mapping of protein conformational features such as folded domains or epitopes.

Some obvious limitations of solution-based labeling include potentially denaturing labeling conditions, such as low pH or sparingly soluble reagents that require addition of organic solvents.[137] Another disadvantages is that derivatization reactions can take anywhere from a few minutes to over 24 hours. Long labeling times could allow for protein conformations to change as residues are progressively modified. Also, side reactions can occur with less specific derivatization reagents, for example, the anhydrides used for acetylation of K also react with tyrosine, which can make it complicated to distinguish (by mass) between isobaric peptide segments. Proteins experience chemical and physical changes e.g. charge-reversal of K residues labeled at neutral pH or the introduction of bulky modifications which can alter structure. Another solution-based labeling method, which has none of these disadvantages is amide backbone hydrogen exchange (HX). HX takes advantage of the natural protein-solvent

interchange of hydrogens, to label proteins with deuterium (^2H) in D_2O solvent. The fundamentals and advantages of HX labeling will be discussed in section *1.5.1*.

1.4.4 General advantages of MS over spectroscopic methods

Before discussing the HX process, some general advantages of MS over CD, FRET and NMR spectroscopy will be outlined. Firstly, MS can be used with relatively impure samples and does not require isolation of the protein of interest. Unlike CD, which averages the secondary structure content of protein molecules in an ensembles or in a mixture, MS spectra give distinct resolvable peaks per protein or protein conformer.[126] While NMR can provide atomic level resolution, it requires more careful sample preparation because signal-overlap can be problematic in protein mixtures. The ability of MS analysis to resolve many protein species in a single sample make it the technique of choice for proteomics-scale analyses (which necessarily include many IDPs).[145] Secondly, MS requires picomole (or even subpicomole) amounts of protein[146-148], compared to NMR which requires micromoles (several orders of magnitude more) of proteins. Requiring low proteins concentrations can be useful for studying IDPs, which can be prone to aggregation at higher concentrations.[149] Other advantages over NMR include ease of analyzing large proteins e.g. hundreds of kiloDalton viral protein complexes[150], much simpler data analysis i.e. relatively simple mass measurements compared to chemical shift assignments. Furthermore, MS with enzymatic digestion, electron transfer dissociation (ETD) or electron-capture dissociation (ECD) provides peptide- to resides-level resolution, which CD and FRET cannot provide.

1.5 Hydrogen exchange mass spectrometry

1.5.1 Fundamentals of protein HX: Linderstøm-Lang theory

Backbone amide hydrogen exchange (HX) is the exchange of protium (^{1}H or simply H) with deuterium (^{2}H or D) which can be measured by MS[151] and NMR[152]. Protein hydrogens are found in backbone and side-chain amides, other side chains (e.g. amines and hydroxyls), and in aliphatic chains. Only backbone amide HX is measured by HX-MS because aliphatic hydrogens do not exchange during labeling and side chains back-exchange (D-H) rapidly in H_2O-based solvent during LC-MS analysis. When a protein is placed in D_2O, HX proceeds as described by the Linderstrøm-Lang model[153], depicted in scheme 1:

$$\text{NH (closed)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{NH (open)} \xrightarrow[D_2O]{k_{HX}} \text{ND (open)} \underset{k_{op}}{\overset{k_{cl}}{\rightleftharpoons}} \text{ND (closed)}$$

Scheme 1

where k_{op} and k_{cl} are rate constants for protein unfolding (i.e. opening) and folding (i.e. closing) respectively; and k_{ch} is the rate constant for hydrogen-deuterium exchange (i.e. chemical exchange). Amides hydrogens involved in hydrogen bonding or those buried in hydrophobic cores (and therefore not solvent-accessible) are in the *closed* state and are said to be *exchange-incompetent*. As local intramolecular hydrogen bonds break or the protein globally unfolds due to naturally-occurring fluctuations (sometimes called *breathing motions*), amide hydrogens enter the *open* state where they are *exchange-competent*. Protein breathing motions could be altered by ligand binding or by changes in solution conditions, such as pH and temperature. In the open state, HX proceeds at the chemical exchange rate (k_{ch}) (see Figure 1.3).

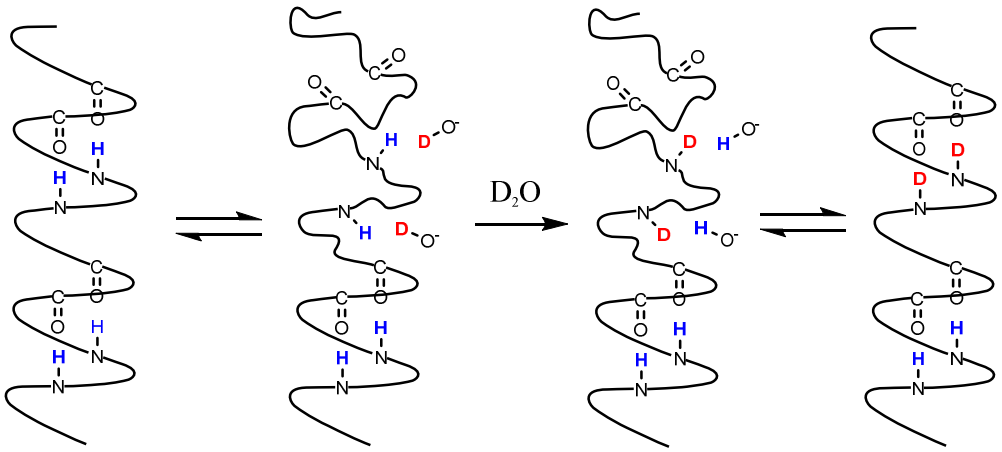


Figure 1.3 HX occurs when intramolecular hydrogen bonds in secondary structure features e.g. α-helices break naturally due to local fluctuations. Hydrogens in intrinsic coil IDPs regions are mostly in the open state compared to those in ordered protein regions. Deutroxide ions are shown because HX is predominantly base-catalyzed at neutral pH.

Chemical exchange is acid- and base-catalyzed and (to a very limited extent) water-catalyzed. Base-catalyzed HX is several orders of magnitude faster than acid-catalyzed HX near neutral pH conditions. The chemical exchange rate constant is described by the following rate constant equation:

$$k_{\mathrm{ch}} = k_{\mathrm{A}}\,[\mathrm{H_3O^+}] + k_{\mathrm{B}}[\mathrm{OH^-}] \qquad (1)$$

where k_{A} and k_{B} rate constants for acid- and base-catalyzed HX respectively. Because water-catalyzed HX can be a trillion times slower than base-catalyzed HX the $k_{\mathrm{w}}[\mathrm{H_2O}]$ term is omitted in equation (1). The HX chemical exchange rate can expressed by equation (2), as follows:

$$\mathrm{rate} = k_{\mathrm{ch}}\,[\mathrm{NH}][\mathrm{OD^-}] = k'[\mathrm{NH}] \qquad (2)$$

HX is a pseudo first-order reaction in buffered solutions, because the concentration of deutroxide ($\mathrm{OD^-}$) ions remains constant, therefore $k_{\mathrm{ch}}[\mathrm{OD^-}] = k'$. Besides pH, other factors affecting chemical exchange rates are temperature, neighboring side chains[154], salt content, and the presence of some denaturants, such as urea.[155]

The chemical exchange rate is a reference rate, representing HX of amides in completely unstructured peptides. The experimentally-observed HX rate (k_{HX}) depends on local or global protein folding, which can modulate intramolecular hydrogen bonds or reduce solvent accessibility of residues; k_{HX} is described by the following equation:

$$k_{\mathrm{HX}} = \frac{k_{\mathrm{op}}}{k_{\mathrm{cl}}} k_{\mathrm{ch}} \qquad (3)$$

under typical HX kinetic limits, also called the EX2 limit, $k_{\mathrm{cl}} \gg k_{\mathrm{ch}}$. In the EX2 limit, opening and closing occurs many times before HX happens, representing the commonly observed HX phenomenon. Conveniently, the ratio $\frac{k_{\mathrm{op}}}{k_{\mathrm{cl}}}$ from equation (3) can be inverted to $\frac{k_{\mathrm{cl}}}{k_{\mathrm{op}}}$ to describe a *protection factor* (PF), which is a measure of protein stability (i.e. protections from HX). For very stable proteins PF $\gg 1$ reflecting that $k_{\mathrm{cl}} \gg k_{\mathrm{op}}$, resulting in slowed HX rates (i.e. $k_{\mathrm{HX}} \ll k_{\mathrm{ch}}$). Furthermore, this relationship fits directly into the expression for the unfolding free energy (ΔG_{U}), in the expression:

$$\Delta G_{\mathrm{U}} = RT \ln \frac{k_{\mathrm{cl}}}{k_{\mathrm{op}}} = RT \ln \mathrm{PF} \qquad (4)$$

where R and T are the gas constant and absolute temperature respectively. In principle, IDPs have a greater preference for the unfolded state than structured proteins, consistent with smaller PF (i.e. less protection from HX) and negative ΔG_{U} values.

The HX kinetics described above are for individual amides, but HX-MS typically yields peptide-level resolution i.e. 3 or more residues. The observed HX kinetics for peptide segments are averaged over the residues in the peptide. Determining the HX kinetics of individual residues within a peptide is an underdetermined problem, i.e. many different sets of rate constants

describe the data equally well, so there is no unique solution. Discussions herein are only concerned with comparisons between protein states, but the governing principles in each HX experiment are similar to those described above.

1.5.2 Advantages of HX-MS

Despite the high cost of D_2O, HX has numerous advantages over other labeling techniques. Firstly, HX is a universal reaction i.e. all proteins in aqueous solvent naturally undergo HX. Therefore, labeling does not involve the addition of reagents or changes in solutions conditions which could alter protein structure. Secondly, HX is an example of *isotopic labeling* instead of residue modification; deuterons and protons have similar physico-chemical properties and are miniscule compared to proteins. Although the strengths of hydrogen and deuterium bonds are different, the overall effect of deuterium uptake on protein structure is assumed (but not confirmed) to be negligible. Moreover, for experiments comparing protein structures in different conditions, only the HX rates matter (and not the physical effects which could result from HX). Also, D_2O and H_2O buffers have slightly different pH, but correcting for this *isotope effect* simply involves a downward shift of 0.4 pH units compared to pH in H_2O buffers.[156] Thirdly, HX occurs under all pH conditions used in protein analysis. Therefore HX does not require specific, potentially structure-altering, pH or temperature conditions and can be measured at various pH/temperature values of interest. For example, the experiments described in chapter 4 of this dissertation were performed at pH 5.5 to slow HX and thereby capture HX kinetics of a random coil IDP. Furthermore, HX is easily controlled by altering pH and temperature meaning that labelled protein can be stored, typically in pH 2.5 and −80 °C conditions, for months before analysis. HX rates increase tenfold with every pH unit and with every 22 °C. Finally, HX can be performed in various aqueous solution conditions e.g. crowded

conditions without altering chemical exchange rates.[157] This characteristic is key for the experiments described in chapters 3 and 4 of this dissertation, which were performed in up to 400 g L^{-1} of polymer crowder. A few other examples of HX performed in complex solution matrixes are also listed in chapter 3.

1.5.3 Hydrogen exchange mass spectrometry workflow

HX-MS has developed into a useful method for probing protein conformation and dynamics in solution (and in the gas phase).[158-160] HX-MS now finds widespread applications in academia and in industry.[161] The HX-MS techniques outlined in this dissertation involves continuous labeling experiments, wherein protein HX kinetics are measured as a function of D_2O exposure time. IN each experiment, HX kinetics are compared to determine conformational effects under two different conditions. For example, changes in the intrinsic coil structure of CaN-RD-AID when it binds to calmodulin (chapter 2) or changes in IDP transient structures in the presence of a polymer crowder (chapters 3 and 4). Slower HX, i.e. protection from HX, generally indicates stabilization of structure in a given protein region. For example, HX was slowed in parts of the CaN-RD after binding calmodulin compared to free CaN-RD (discussed in chapter 2). Faster HX, i.e. deprotection form HX, indicates destabilization of protein structure. For example, HX was faster in transiently-helical regions of ACTR, suggesting that 300 g L^{-1} Ficoll destabilized transient α-helices. The steps involved in typical continuous labeling HX-MS experiments are shown in Figure 1.4 and are as follows:

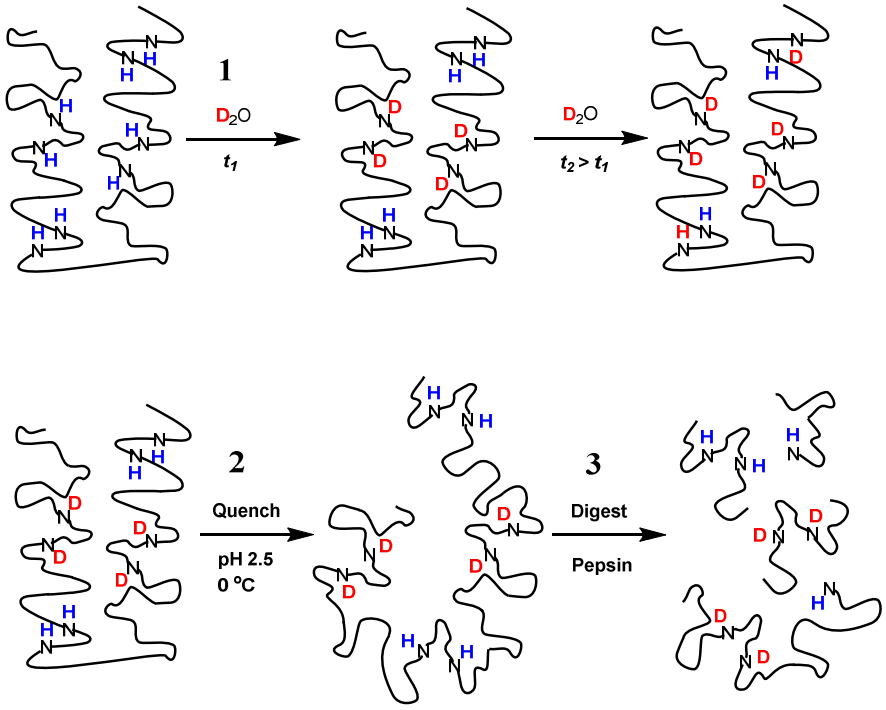


Figure 1.4 Flexible, unstructured regions exchange faster than rigid, structured regions. Proteins are exposed to deuterium for time, *t.* After digestion, deuterated peptides are desalted on a trap, separated on an analytical column, and detected by ESI-MS.

In step 1, protein is labeled by diluting the sample, typically by factors of 2-20, with a D_2O buffer of equivalent pH (usually reported as pD indicating correction for the isotope effect[156]). The protein is labeled for different amounts of time, typically seconds to hours, in the D_2O buffer. Experiments are usually performed in triplicate for each time-point. For HX in crowded conditions, as described in chapters 3 and 4, labeling and sample buffers contains equivalent amounts of crowder (Ficoll) to avoid potentially altering protein structure due to dilution effects. Ficoll contains labile hydroxyl groups, so it was pre-deuterated to ensure that equivalent amounts of deuterium were available for HX in both (crowded and uncrowded) conditions. Pre-labeling is not necessary for crowders (e.g. PVP) which do not contain labile hydrogens. For comparison of free and ligand-bound states, as described in chapter 2, the ligand-bound protein must reach equilibrium before labeling to ensure that observed exchange (k_{HX}) is predominantly due to conformations in the *fully-bound* state, and not due to intermediate states.

These factors are important to ensure that any changes in HX kinetics can be attributed only to specific experimental variables and not to changes in protein equilibria during or after dilution in D_2O buffer. Some denaturants, such as urea, alters k_{ch} and so should be avoided comparative HX experiments if one of the conditions excludes the denaturant.[155]

During labeling, flexible regions, which are less protected from HX will fully exchange during the shorter labeling time-points i.e. milliseconds to seconds. Intrinsic coil IDPs exchange rapidly, and all residues can be fully deuterated within a few seconds, in neutral pH at room temperature. Less flexible regions, involved in intramolecular hydrogen bonds or buried in hydrophobic cores take much longer to exchange i.e. hours, weeks or even years. For example, some residues of native cytochrome c exchange have protection factors of about 10^6, representing a million-fold slower HX compared to unfolded cytochrome c.[162] Differences in HX protection factors can be used to distinguish between structured and unstructured regions in a protein. When different conditions substantially alter HX rates in a particular protein segments e.g. calmodulin binding in chapter 2, HX can be performed near neutral pH and room temperature. However, IDP labeling could be performed at one or two pH units below pH 7, especially if the IDP exchanges very rapidly at every residue or if the anticipated HX differences in different conditions are subtle. This is done to slow the HX kinetics so that an increase in uptake is observed with as many time increments as possible.

After each labeling time, the reaction is quenched by diluting, typically 1:1, with pH 2.5 *quench buffer* and chilling to 0 °C (step 2 of Figure 1.4). At pH 2.5, backbone amides HX is at its lowest rate[163] and chilled conditions further slow HX, reproducibly locking deuterium labels in place before MS analysis. To ensure that the reaction mixture is brought to pH 2.5, the quench buffer is usually a strong acid e.g. HCl or has higher buffering capacity than the labeling buffer

e.g. 200 mM phosphate (pH 2.5) quench buffer compared to 20 mM phosphate (pH 7.0) labeling buffer. The quench buffer is H_2O-based, which can lead to deuterium loss via back-exchange (D-H). However, back-exchange is slowed in the chilled pH 2.5 quench buffer, similarly to how forward exchange (H-D) is slowed. Only side chains exchange rapidly enough to completely lose their deuterons by the time MS measurements can be made. The loss of side-chain deuterons can be advantageous for HX-MS because it removes ambiguity between backbone and side chain HX, which result in the same mass change. Quenched, intact, protein can be analyzed by MS, revealing global changes in protein flexibility brought about by different conditions. In chapter 2, MS measurements of deuterium uptake for intact CaN-RD-AID, in free versus calmodulin-bound forms, is compared.

More often, the quenched protein is digested with an acid-resistant protease, such as pepsin, in solution or immobilized in an LC column[164] (step 3 in Figure 1.4). Digestion localizes deuterium labels along the protein sequence with peptide-level resolution providing information about changes in specific protein segments. CaN-RD-AID was later digested to localize the regions that became structured, as had been suggested by the intact HX measurement. The spatial resolution of HX labeling depends on the lengths of peptide segments obtained after proteolysis. Pepsin is a preferred enzyme for HX because it has less specific cleavages (than trypsin, for example), resulting in numerous, relatively short peptides (typically 5 to 20 residues) with multiple overlapping segments. All these features of peptic peptide maps, greatly improve resolution compared to trypic (or other) digests. As discussed earlier, resolution can be enhanced by using multiple enzymes together, or in separate experiments then combining peptide maps.

After digestion the peptides are captured on a desalting trap where they are concentrated and salts are washed away. Desalting prevents salts from accumulating in the ESI source or the

MS optics, where they can contaminate subsequent sample runs (with salt-adducts) or reduce instrument sensitivity. Thereafter, peptides are separated on an analytical column, ionized by an ESI source and detected by MS. Separation of peptides improves the resolution of different peptide segments by reducing ionization suppression from large quantities of co-eluting peptides. HPLC separation also aids in data analysis because the same peptide typically elutes within the same narrow (<30 s) retention time window. Finally, separation, usually achieved over a short gradient elution (typically about 5 minutes), ensures cleaner peptide spectra with reduced overlap in mass-to-charge space. Clean spectra are crucial for finding HX spectral features used to determine deuterium uptake.

1.5.4 Automated HX-MS

The HX-MS experiments described in chapter 2 were performed manually. Experiments in chapters 3 and 4 were automated using LEAP HD/X PAL robotic system (LEAP Technologies, Carborro, NC, USA). While the principles governing the HX labeling remain unchanged, automation improves efficiency by performing replicates faster. For example, two replicates were performed for each time point separated by a few weeks for the study described in chapter 2. In contrast, triplicate data was acquired for similar time points in about 36 hours of total experiment time. Automation also eliminates the reliance on human timing and pipetting skills, thereby reducing random and systematic errors. Furthermore, the crowding experiments described in chapters 3 and 4 would have been impractical without automated timing of several LC injections and valve switches. Despite the advantages of automation, manual HX training can be beneficial to better understand the principles underlying HX workflows. Understanding HX workflows could help an experimenter to develop complicated HX experiments, such as the SCX-based cleanup outlined in chapter 3.

1.5.5 Data analysis

Initially, undeuterated peptic peptides are mapped onto the protein sequence by accurate mass determination, and confirmed by MS/MS using collision-induced dissociation or other fragmentation methods. Under the EX2 limit, described by equation (3) in section *1.5.1*, peptide HX spectra appear as distinct isotope clusters of increasing mass (depicted in Figure 1.5). The isotope distribution of undeuterated peptides reflects the natural abundance of isotopes, mainly ^{12}C, ^{13}C and ^{15}N. The peaks distribution changes after deuteration because individual protein molecules, in an ensemble, incorporate different amounts of deuterium during labeling, resulting in a spectral envelope resembling a binomial distribution.[165]

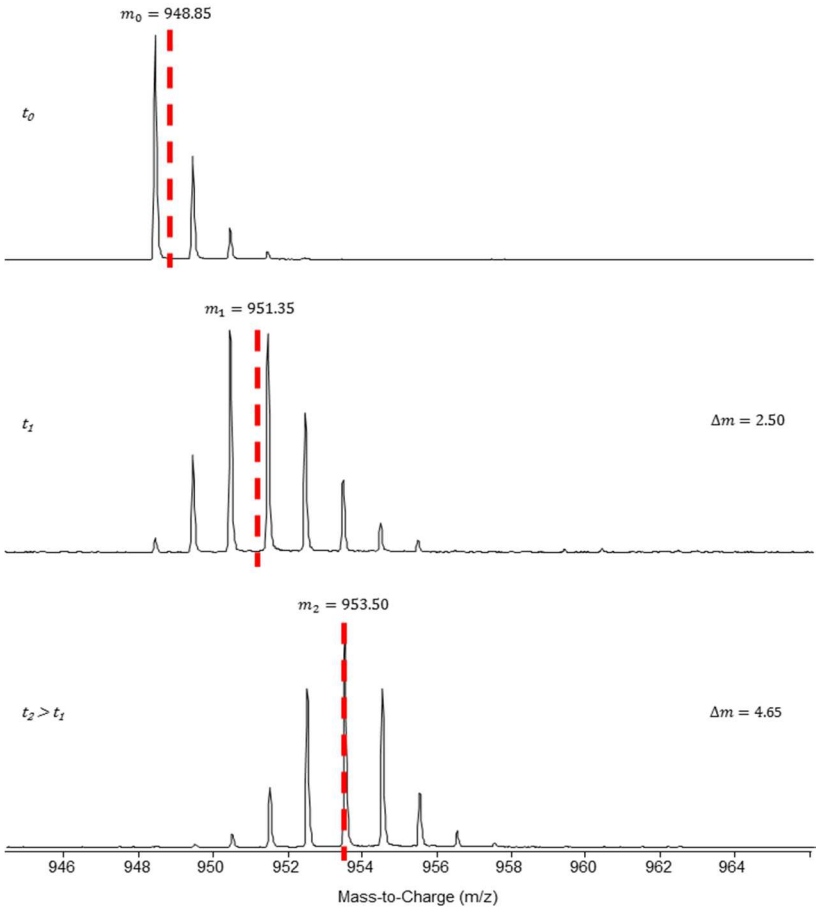


Figure 1.5 Peptide HX spectra are distinct isotope clusters under EX2 limits. Mass increase is calculated using the centroid of each isotope cluster. t represents D2O exposure time, with t0 denoting spectra acquired in an udeuterated control experiment with H2O, instead of D2O, labeling buffer.

Free, open-source and commercially-available software packages can be used to process HX-MS data sets. Software package identify HX spectral features, matching them with the corresponding undeuterated peptide spectra, and compute the average deuterated mass of each peptide. Average deuterated mass is calculated from the weighted average of peak intensities and mass-to-charge (m/z) values, using the center of mass of isotope clusters, at each time point. The mass increase (Δm), resulting from deuterium uptake is determined for each peptide as follows:

$$\Delta m = \bar{m}_{\mathrm{t,c}} - m_0 \qquad (5)$$

where $\bar{m}_{\mathrm{t,c}}$ and m_0 represent average deuterated mass from triplicate measurements, after D_2O exposure for time (t) under condition (c), and the peptide's theoretical average mass (at t = 0) respectively. Uptake curves are constructed by plotting Δm as a function of D_2O exposure time. Deuterium uptake in different conditions is compared by plotting the respective uptake curves on the same graph. Comparatively large HX differences are directly discernable from the differential uptake curves. HX data presented in chapter 2 shows that free CaN-RD-AID was completely deuterated at every residue after 5 s of labeling, while peptides from the calmodulin-binding domain were not completely deuterated, even after 24 hours of labeling, in pH 7 at room temperature. In such instances, conclusions are drawn by merely inspecting the uptake curves. However, the crowder-induced changes presented in chapters 3 and 4 are comparatively small and the curves from crowded and uncrowded conditions are almost overlapping. In such cases, further statistical analysis and other mathematical approaches, described in both chapters, can be required to assign HX differences.

1.5.6 HX-MS of proteins in crowded samples

Typical HX-MS workflows only include a desalting step before LC separation and MS detection, as described in section *1.5.3*. Such a setup is impractical for highly concentrated polymer solutions (required to simulate intracellular conditions). Crowded solutions, containing 300 g L^{-1} and 400 g L^{-1} of Ficoll, are described in chapters 3 and 4. These solutions are very viscous compared to dilute samples, so the pressure limits of some LC pumps could be exceeded, especially under the chilled (quench) conditions inside narrow-bore LC-lines. Also, the solubility of polymers should decrease in the chilled LC solvents, which could result in the accumulation of undissolved polymers in LC lines and columns. Such accumulation would also increase the pressure in LC lines over time, with consecutive crowded experimental runs. Moreover, accumulation of polymers in pepsin, desalting and analytical columns could reduce the efficiency of digestion, trapping and separation, respectively. To avoid these problems, intervening cleaning steps, such as those described in chapter 3, can be used. These cleaning steps are particularly important to ensure uninterrupted LC flow during consecutive HX-MS experiments, sometimes programmed to run (unattended) for over 24 hrs.

Another problem, arising from the high crowder concentrations, about a thousand-fold in excess of protein, is considerable interference with peptide MS detection. Co-eluting crowders (and other contaminants from polymers manufacturing and packaging) would interfere with peptide spectra, impacting the ability of software to resolve peptide isotope clusters. Resolving peptide isotope clusters is especially important for HX feature finding and centroid calculations, necessary to determine deuterium uptake. Besides interfering with peptide spectra, polymers could contaminate the instrument components and reduce sensitivity over time or cause perpetual interference. Additionally, the desalting step of typical HX workflows could be inadequately to

remove the crowder because of their quantities, and because their long polymers chains could bind strongly to reversed-phase columns. To counter these problems, additional columns can be added to HX workflows to separate crowder and peptides, based on physico-chemical differences between peptides and crowders. In principle, any separation mode, which segregates either protein or crowder under quench conditions, can be incorporated into an HX-MS workflow. For example, strong anion exchange has been used to extract highly concentrated nucleic acids from proteins (under quench conditions).[131] In addition to extracting the peptides, further wash steps could be required to remove residual contaminants and to regenerate columns between experimental runs. Some consideration is required in method development to prevent harsh wash solutions, such as organic solvents, from damaging sensitive columns materials, such as immobilized pepsin. The experiments described in chapters 3 and 4 employ combinations of injections, pumped solvents, multiple serial valves, and timed valve switching to perform different column washing and regeneration steps. Figure 1.6 shows a comparison between a

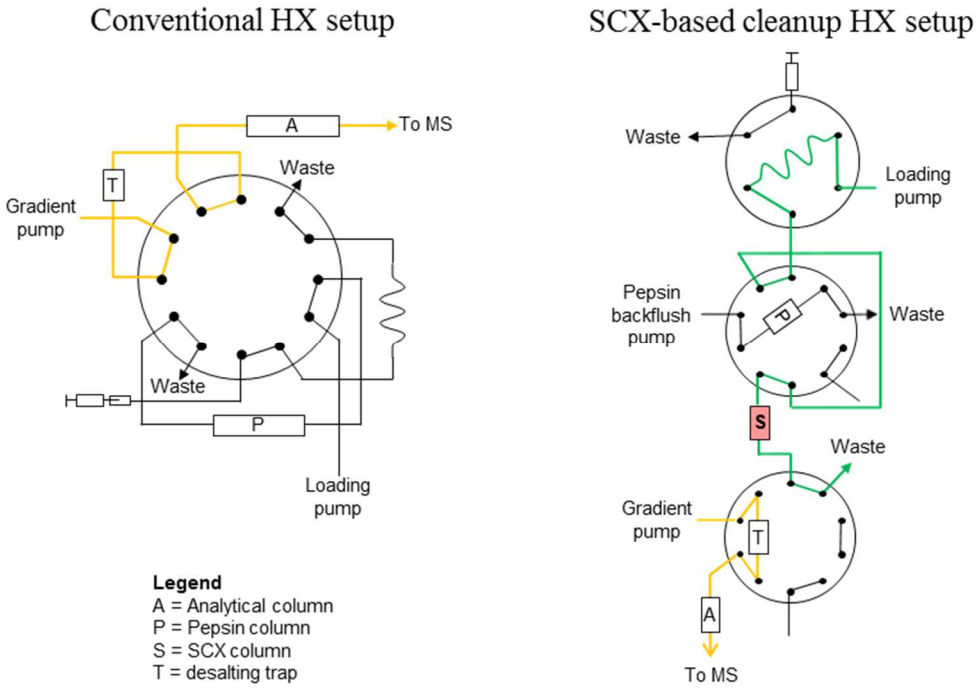


Figure 1.6 The SCX-based cleanup has multiple serial valves and an SCX column, unlike conventional the HX setup. Most peptides are positively-charged in quench conditions, pH 2.5, and are trapped on the SCX column, while neutral Ficoll flows to waste. Valve switches separate SCX and pepsin column during SCX washing, with methanol. In the position shown, gradient elution (orange flow-path) and SCX regeneration (green flow-path) can occur simultaneously.

conventional HX-MS setup, as described in chapter 2, and the SCX-based setup for removal of polymer crowders, described in chapters 3 and 4. Using the multiple valves, timed to switch automatically, columns can be isolated during different wash steps. Further columns can be washed as peptides are detected by MS, simultaneously.

1.6 References

[1] Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis, *Nature 181*, 662-666.

[2] Perutz, M. (1960) Structure of hemoglobin, In *Brookhaven. Symp. Biol.*, pp 165-183.

[3] Anfinsen, C. B., Haber, E., Sela, M., and White, F. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, *Proc. Natl. Acad. Sci. U.S.A. 47*, 1309-1314.

[4] Huber, R., and Bennett, W. S. (1983) Functional significance of flexibility in proteins, *Biopolymers 22*, 261-279.

[5] Sigler, P. B. (1988) Transcriptional activation. Acid blobs and negative noodles, *Nature 333*, 210.

[6] Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, *J. Mol. Biol. 293*, 321-331.

[7] Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat. rev. Mol. cell biol. 6*, 197-208.

[8] Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) Intrinsic disorder in transcription factors, *Biochemistry 45*, 6873-6888.

[9] Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) Function and structure of inherently disordered proteins, *Curr. Opin. Struct. Biol. 18*, 756-764.

[10] Dunker, A. K., and Obradovic, Z. (2001) The protein trinity—linking function and disorder, *Nat. Biotechnol. 19*, 805-806.

[11] Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci. 11*, 739-756.

[12] Hilser, V. J., García-Moreno E, B., Oas, T. G., Kapp, G., and Whitten, S. T. (2006) A statistical thermodynamic model of the protein ensemble, *Chem. Rev. 106*, 1545-1558.

[13] Ohgushi, M., and Wada, A. (1983) 'Molten□globule state': a compact form of globular proteins with mobile side□chains, *FEBS letters 164*, 21-24.

[14] Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators, *Nature 415*, 549-553.

[15] Ebert, M.-O., Bae, S.-H., Dyson, H. J., and Wright, P. E. (2008) NMR relaxation study of the complex formed between CBP and the cctivation domain of the nuclear hormone receptor coactivator ACTR†, *Biochemistry 47*, 1299-1308.

[16] Chaffotte, A. F., Guijarro, J. I., Guillou, Y., Delepierre, M., and Goldberg, M. E. (1997) The "pre-molten globule," a new intermediate in protein folding, *J. Prot. Chem. 16*, 433-439.

[17] Uversky, V. N., and Ptitsyn, O. B. (1996) Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature, *J. Mol. Biol. 255*, 215-228.

[18] Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N., and Obradovic, Z. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions, *J. Proteom. Res. 6*, 1882-1898.

[19] Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., and Sosnick, T. R. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12491-12496.

[20] Ferron, F., Longhi, S., Canard, B., and Karlin, D. (2006) A practical overview of protein disorder prediction methods, *Prot. 65*, 1-14.

[21] Das, R. K., and Pappu, R. V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, *Proc. Natl. Acad. Sci. U.S.A. 110*, 13392-13397.

[22] Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., and Pappu, R. V. (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins, *Proc. Natl. Acad. Sci. U.S.A. 107*, 8183-8188.

[23] Müller-Späth, S., Soranno, A., Hirschfeld, V., Hofmann, H., Rüegger, S., Reymond, L., Nettels, D., and Schuler, B. (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins, *Proc. Natl. Acad. Sci. U.S.A. 107*, 14609-14614.

[24] Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations, In *Pac. Symp. Biocomput.*, pp 473-484.

[25] Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998) Thousands of proteins likely to have long disordered regions, In *Pac. Symp. Biocomput.*, pp 437-448.

[26] Mathura, V., Braun, W., Garner, E., Young, J., Takayama, S., Brown, C., and Dunker, A. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder, In *Pac. Symp. Biocomput.*, pp 89-100.

[27] He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., and Dunker, A. K. (2009) Predicting intrinsic disorder in proteins: an overview, *Cell Res. 19*, 929-949.

[28] Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005) Comparing and combining predictors of mostly disordered proteins, *Biochemistry 44*, 1989-2000.

[29] Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting protein disorder for N-, C-and internal regions, *Genome Inform. Ser. 10*, 30-40.

[30] Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005) FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded, *Bioinformatics 21*, 3435-3438.

[31] Dunker, A. K., Romero, P., Obradovic, Z., Garner, E. C., and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes, *Genome Inform. Ser. 11*, 161-171.

[32] Tompa, P., and Varadi, M. (2014) Predicting the predictive power of IDP ensembles, *Structure 22*, 177-178.

[33] Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J. Mol. Biol. 337*, 635-645.

[34] van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014) Classification of intrinsically disordered regions and proteins, *Chem. Rev. 114*, 6589-6631.

[35] Uversky, V. N., and Dunker, A. K. (2010) Understanding protein non-folding, *Biochim. Biophys. Acta-Prot. Proteom. 1804*, 1231-1264.

[36] Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins: Struct., Funct., Bioinf. 42*, 38-48.

[37] Burger, V. M., Nolasco, D. O., and Stultz, C. M. (2016) Expanding the range of protein function at the far end of the order-structure continuum, *J. Biol. Chem. 291*, 6706-6713.

[38] Gsponer, J., and Babu, M. M. (2009) The rules of disorder or why disorder rules, *Prog. Biophys. Mol. Biol. 99*, 94-103.

[39] Babu, M. M., van der Lee, R., de Groot, N. S., and Gsponer, J. (2011) Intrinsically disordered proteins: regulation and disease, *Curr. Opin. Struct. Biol. 21*, 432-440.

[40] Daughdrill, G. W., Narayanaswami, P., Gilmore, S. H., Belczyk, A., and Brown, C. J. (2007) Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation, *J. Mol. Evol. 65*, 277-288.

[41] Chen, H., Lin, R. J., Schiltz, R. L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M. L., Nakatani, Y., and Evans, R. M. (1997) Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300, *Cell 90*, 569-580.

[42] Shikama, N., Chan, H. M., Krstic-Demonacos, M., Smith, L., Lee, C.-W., Cairns, W., and La Thangue, N. B. (2000) Functional interaction between nucleosome assembly proteins and p300/CREB-binding protein family coactivators, *Mol. Cell. Biol. 20*, 8933-8943.

[43] Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., and Dunker, A. K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins, *J. Mol. Biol. 323*, 573-584.

[44] Uversky, V. N., Davé, V., Iakoucheva, L. M., Malaney, P., Metallo, S. J., Pathak, R. R., and Joerger, A. C. (2014) Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases, *Chem. Rev. 114*, 6844-6879.

[45] Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept, *Annu. Rev. Biophys. 37*, 215-246.

[46] Sunde, M., McGrath, K. C., Young, L., Matthews, J. M., Chua, E. L., Mackay, J. P., and Death, A. K. (2004) TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer, *Cancer Res. 64*, 2766-2773.

[47] Glenner, G. G., and Wong, C. W. (1984) Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein, *Biochem. Biophys. Res. Commun. 122*, 1131-1135.

[48] Dev, K. K., van der Putten, H., Sommer, B., and Rovelli, G. (2003) Part I: parkin-associated proteins and Parkinson’s disease, *Neuropharmacology 45*, 1-13.

[49] Molkentin, J. D., Lu, J.-R., Antos, C. L., Markham, B., Richardson, J., Robbins, J., Grant, S. R., and Olson, E. N. (1998) A calcineurin-dependent transcriptional pathway for cardiac hypertrophy, *Cell 93*, 215-228.

[50] Heisel, O., Heisel, R., Balshaw, R., and Keown, P. (2004) New onset diabetes mellitus in patients receiving calcineurin inhibitors: a systematic review and meta□analysis, *Am. J. Transplant. 4*, 583-595.

[51] Miyakawa, T., Leiter, L. M., Gerber, D. J., Gainetdinov, R. R., Sotnikova, T. D., Zeng, H., Caron, M. G., and Tonegawa, S. (2003) Conditional calcineurin knockout mice exhibit multiple abnormal behaviors related to schizophrenia, *Proc. Natl. Acad. Sci. U.S.A. 100*, 8987-8992.

[52] Cheng, Y., LeGall, T., Oldfield, C. J., Mueller, J. P., Van, Y.-Y. J., Romero, P., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2006) Rational drug design via intrinsically disordered protein, *Trends Biotechnol. 24*, 435-442.

[53] Uversky, V. N. (2010) Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D2 concept, *Expert Rev. Proteomics 7*, 543-564.

[54] Tompa, P. (2012) Intrinsically disordered proteins: a 10-year recap, *Trends Biochem. Sci. 37*, 509-516.

[55] Arkin, M. R., Tang, Y., and Wells, J. A. (2014) Small-molecule inhibitors of protein-protein interactions: progressing toward the reality, *Chem. Biol. 21*, 1102-1114.

[56] Parthasarathi, L., Casey, F., Stein, A., Aloy, P., and Shields, D. C. (2008) Approved drug mimics of short peptide ligands from protein interaction motifs, *J. Chem. Inf. Model.48*, 1943-1948.

[57] Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., and Moomaw, E. W. (1995) Crystal

structures of human calcineurin and the human FKBP12–FK506–calcineurin complex, *Nature 378*, 641–644

[58] Fulton, A. B. (1982) How crowded is the cytoplasm? Minireviews, *Cell 30*, 345-347.

[59] Gershon, N. D., Porter, K. R., and Trus, B. L. (1985) The cytoplasmic matrix: its volume and surface area and the diffusion of molecules through it, *Proc. Natl. Acad. Sci. U.S.A. 82*, 5030-5034.

[60] Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli, *J. Mol. Biol. 222*, 599-620.

[61] Zimmerman, S. B., and Minton, A. P. (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences, *Annu. Rev. Biophys. Biomol. Struct. 22*, 27-65.

[62] Minton, A. P. (1983) The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences, *Mol. Cell. Biochem. 55*, 119-140.

[63] Zhou, H.-X., Rivas, G., and Minton, A. P. (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences, *Annu. Rev. Biophys. 37*, 375-397.

[64] Elcock, A. H. (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment, *Curr. Opin. Struct. Biol. 20*, 196-206.

[65] Sarkar, M., Lu, J., and Pielak, G. J. (2014) Protein crowder charge and protein stability, *Biochemistry 53*, 1601-1606.

[66] Ogston, A., and Phelps, C. (1961) The partition of solutes between buffer solutions and solutions containing hyaluronic acid, *Biochem. J. 78*, 827.

[67] Ogston, A., and Sherman, T. (1961) Effects of hyaluronic acid upon diffusion of solutes and flow of solvent, *J. Physiol. 156*, 67.

[68] Laurent, T., and Ogston, A. (1963) The interaction between polysaccharides and other macromolecules. 4. The osmotic pressure of mixtures of serum albumin and hyaluronic acid, *Biochem. J. 89*, 249.

[69] Minton, A., and Wilf, J. (1981) Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase, *Biochemistry 20*, 4821-4826.

[70] Minton, A. P. (1981) Excluded volume as a determinant of macromolecular structure and reactivity, *Biopolymers 20*, 2093-2120.

[71] Rohwer, J. M., Postma, P. W., Kholodenko, B. N., and Westerhoff, H. V. (1998) Implications of macromolecular crowding for signal transduction and metabolite channeling, *Proc. Natl. Acad. Sci. U.S.A. 95*, 10547-10552.

[72] Dhar, A., Samiotakis, A., Ebbinghaus, S., Nienhaus, L., Homouz, D., Gruebele, M., and Cheung, M. S. (2010) Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding, *Proc. Natl. Acad. Sci. U.S.A. 107*, 17586-17591.

[73] Martin, J., and Hartl, F.-U. (1997) The effect of macromolecular crowding on chaperonin-mediated protein folding, *Proc. Natl. Acad. Sci. U.S.A. 94*, 1107-1112.

[74] Ellis, R. J. (1997) Molecular chaperones: avoiding the crowd, *Curr. Biol. 7*, R531-R533.

[75] Serber, Z., Keatinge-Clay, A. T., Ledwidge, R., Kelly, A. E., Miller, S. M., and Dötsch, V. (2001) High-resolution macromolecular NMR spectroscopy inside living cells, *J. Am. Chem. Soc. 123*, 2446-2447.

[76] Hall, D., and Minton, A. P. (2003) Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges, *Biochim. Biophys. Acta 1649*, 127-139.

[77] McNulty, B. C., Young, G. B., and Pielak, G. J. (2006) Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder, *J. Mol. Biol. 355*, 893-897.

[78] Li, C., Charlton, L. M., Lakkavaram, A., Seagle, C., Wang, G., Young, G. B., Macdonald, J. M., and Pielak, G. J. (2008) Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: Implications for in-cell NMR spectroscopy, *J. Am. Chem. Soc. 130*, 6310-6311.

[79] Minton, A. P. (2005) Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: macromolecular crowding and protein stability revisited, *Biophys. J. 88*, 971-985.

[80] Zhou, H.-X., and Dill, K. A. (2001) Stabilization of proteins in confined spaces, *Biochemistry 40*, 11289-11293.

[81] Miklos, A. C., Li, C., Sharaf, N. G., and Pielak, G. J. (2010) Volume exclusion and soft interaction effects on protein stability under crowded conditions, *Biochemistry 49*, 6984-6991.

[82] Miklos, A. C., Sarkar, M., Wang, Y., and Pielak, G. J. (2011) Protein crowding tunes protein stability, *J. Am. Chem. Soc. 133*, 7116-7120.

[83] Mukherjee, S., Waegele, M. M., Chowdhury, P., Guo, L., and Gai, F. (2009) Effect of macromolecular crowding on protein folding dynamics at the secondary structure level, *J. Mol. Biol. 393*, 227-236.

[84] Soranno, A., Koenig, I., Borgia, M. B., Hofmann, H., Zosel, F., Nettels, D., and Schuler, B. (2014) Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments, *Proc. Natl. Acad. Sci. U.S.A. 111*, 4874-4879.

[85] de Gennes, P.-G. (1979) *Scaling concepts in polymer physics*, Cornell University Press, Ithaca, NY.

[86] Monteith, W. B., Cohen, R. D., Smith, A. E., Guzman-Cisneros, E., and Pielak, G. J. (2015) Quinary structure modulates protein stability in cells, *Proc. Natl. Acad. Sci. U.S.A. 112*, 1739-1742.

[87] Spillantini, M. G., Schmidt, M. L., Lee, V. M.-Y., Trojanowski, J. Q., Jakes, R., and Goedert, M. (1997) α-Synuclein in Lewy bodies, *Nature 388*, 839-840.

[88] Keppel, T. R., Howard, B. A., and Weis, D. D. (2011) Mapping unstructured regions and synergistic folding in intrinsically disordered proteins with amide H/D exchange mass spectrometry, *Biochemistry 50*, 8722-8732.

[89] Kjaergaard, M., Teilum, K., and Poulsen, F. M. (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP, *Proc. Natl. Acad. Sci. U.S.A. 107*, 12535-12540.

[90] Kjaergaard, M., Nørholm, A. B., Hendus–Altenburger, R., Pedersen, S. F., Poulsen, F. M., and Kragelund, B. B. (2010) Temperature□dependent structural changes in intrinsically disordered proteins: Formation of α–helices or loss of polyproline II?, *Protein Sci. 19*, 1555-1564.

[91] Keppel, T. R., and Weis, D. D. (2015) Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging, *J. Am. Soc. Mass. Spectrom. 26*, 547-554.

[92] Semisotnov, G., Rodionova, N., Razgulyaev, O., Uversky, V., Gripas, A., and Gilmanshin, R. (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe, *Biopolymers 31*, 119-128.

[93] Fontana, A., de Laureto, P. P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) Probing the partly folded states of proteins by limited proteolysis, *Folding and Design 2*, R17-R26.

[94] Receveur□Bréchot, V., Bourhis, J. M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding, *Proteins: Struct., Funct., Bioinf. 62*, 24-45.

[95] Mendoza, C., Figueirido, F., and Tasayco, M. L. (2003) DSC studies of a family of natively disordered fragments from Escherichia coli thioredoxin: surface burial in intrinsic coils, *Biochemistry 42*, 3349-3358.

[96] Uversky, V. N., Li, J., and Fink, A. L. (2001) Evidence for a partially folded intermediate in α-synuclein fibril formation, *J. Biol. Chem. 276*, 10737-10744

[97] Sibille, N., and Bernadó, P. (2012) Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS, *Biochem. Soc. Trans. 40*, 955-962.

[98] Rhoades, E., Ramlall, T. F., Webb, W. W., and Eliezer, D. (2006) Quantification of α-synuclein binding to lipid vesicles using fluorescence correlation spectroscopy, *Biophys. J. 90*, 4692-4700.

[99] Nettels, D., Müller-Späth, S., Küster, F., Hofmann, H., Haenni, D., Rüegger, S., Reymond, L., Hoffmann, A., Kubelka, J., and Heinz, B. (2009) Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 106*, 20740-20745.

[100] Tompa, P., and Fersht, A. (2009) *Structure and function of intrinsically disordered proteins*, CRC Press, Boca Raton, FL

[101] Gast, K., Damaschun, H., Misselwitz, R., Müller-Frohne, M., Zirwer, D., and Damaschun, G. (1994) Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate, *Eur. Biophys. J. 23*, 297-305.

[102] Greenfield, N. J., and Fasman, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry 8*, 4108-4116.

[103] Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nature Protocols 1*, 2876-2890.

[104] Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism, *Biochim. Biophys. Acta -Prot. Proteom. 1751*, 119-139.

[105] Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, *Biopolymers 89*, 392-400.

[106] Schweers, O., Schönbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994) Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure, *J. Biol. Chem. 269*, 24290-24297.

[107] Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T. (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded, *Biochemistry 35*, 13709-13715.

[108] Koutsioubas, A., Lairez, D., Combet, S., Fadda, G., Longeville, S., and Zalczer, G. (2012) Crowding effect on helix-coil transition: beyond entropic stabilization, *J. Phys. Chem. 136*, 215101-215108.

[109] Flaugh, S. L., and Lumb, K. J. (2001) Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27Kip1, *Biomacromolecules 2*, 538-540.

[110] Clegg, R. M. (2009) Förster resonance energy transfer—FRET what is it, why do it, and how it's done, *Lab. Techiq. Biochem. Mol. Biol. 33*, 1-57.

[111] Jares-Erijman, E. A., and Jovin, T. M. (2003) FRET imaging, *Nat. Biotechnol. 21*, 1387-1395.

[112] Ohashi, T., Galiacy, S. D., Briscoe, G., and Erickson, H. P. (2007) An experimental study of GFP□based FRET, with application to intrinsically unstructured proteins, *Protein Sci. 16*, 1429-1438.

[113] Jensen, M. R., Markwick, P. R., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernado, P., and Blackledge, M. (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings, *Structure 17*, 1169-1185.

[114] Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc. Natl. Acad. Sci. U.S.A. 94*, 12366-12371.

[115] Pelton, J. T., and McLean, L. R. (2000) Spectroscopic methods for analysis of protein secondary structure, *Anal. Biochem. 277*, 167-176.

[116] Kjaergaard, M., Iešmantavičius, V., and Poulsen, F. M. (2011) The interplay between transient α□helix formation and side chain rotamer distributions in disordered proteins probed by methyl chemical shifts, *Protein Sci. 20*, 2023-2034.

[117] Miklos, A. C., Li, C., Sorrell, C. D., Lyon, L. A., and Pielak, G. J. (2011) An upper limit for macromolecular crowding effects, *BMC Biophys. 4*, 13.

[118] Wang, Y., Sarkar, M., Smith, A. E., Krois, A. S., and Pielak, G. J. (2012) Macromolecular crowding and protein stability, *J. Am. Chem. Soc. 134*, 16614-16618.

[119] Dass, C. (2007) *Fundamentals of contemporary mass spectrometry*, Vol. 16, John Wiley & Sons, Hoboken, NJ

[120] Whitehouse, C. M., Dreyer, R., Yamashita, M., and Fenn, J. (1989) Electrospray ionization for mass-spectrometry of large biomolecules, *Science 246*, 64-71.

[121] Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Anal. Chem. 60*, 2299-2301.

[122] Heck, A. J. (2008) Native mass spectrometry: a bridge between interactomics and structural biology, *Nature Methods 5*, 927-933.

[123] Konermann, L., Pan, J., and Liu, Y.-H. (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics, *Chem. Soc. Rev. 40*, 1224-1234.

[124] Natalello, A., Santambrogio, C., and Grandori, R. (2017) Are charge-state distributions a reliable tool describing molecular ensembles of intrinsically disordered proteins by native MS?, *J. Am. Soc. Mass. Spectrom. 28*, 21-28.

[125] D'Urzo, A., Konijnenberg, A., Rossetti, G., Habchi, J., Li, J., Carloni, P., Sobott, F., Longhi, S., and Grandori, R. (2015) Molecular basis for structural heterogeneity of an intrinsically disordered protein bound to a partner by combined ESI-IM-MS and modeling, *J. Am. Soc. Mass. Spectrom. 26*, 472-481.

[126] Frimpong, A. K., Abzalimov, R. R., Uversky, V. N., and Kaltashov, I. A. (2010) Characterization of intrinsically disordered proteins with electrospray ionization mass

spectrometry: Conformational heterogeneity of α□synuclein, *Proteins: Struct., Funct., Bioinf. 78*, 714-722.

[127] Shvartsburg, A. A., Tang, K., and Smith, R. D. (2004) Modeling the resolution and sensitivity of FAIMS analyses, *J. Am. Soc. Mass. Spectrom. 15*, 1487-1498.

[128] Breuker, K., and McLafferty, F. W. (2008) Stepwise evolution of protein native structure with electrospray into the gas phase, 10− 12 to 102 s, *Proc. Natl. Acad. Sci. U.S.A. 105*, 18145-18152.

[129] Hebling, C. M., Morgan, C. R., Stafford, D. W., Jorgenson, J. W., Rand, K. D., and Engen, J. R. (2010) Conformational analysis of membrane proteins in phospholipid bilayer nanodiscs by hydrogen exchange mass spectrometry, *Anal. Chem. 82*, 5415-5419.

[130] Rey, M., Mrázek, H., Pompach, P., Novák, P., Pelosi, L., Brandolin, G., Forest, E., Havlíček, V., and Man, P. (2010) Effective removal of nonionic detergents in protein mass spectrometry, hydrogen/deuterium exchange, and proteomics, *Anal. Chem. 82*, 5107-5116.

[131] Sperry, J. B., Wilcox, J. M., and Gross, M. L. (2008) Strong anion exchange for studying protein–DNA interactions by H/D exchange mass spectrometry, *J. Am. Soc. Mass. Spectrom. 19*, 887-890.

[132] Konermann, L., Tong, X., and Pan, Y. (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches, *J. Mass Spectrom. 43*, 1021-1036.

[133] Kiselar, J. G., and Chance, M. R. (2010) Future directions of structural mass spectrometry using hydroxyl radical footprinting, *J. Mass Spectrom. 45*, 1373-1382.

[134] Wang, L., and Chance, M. R. (2011) Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting, *Anal. Chem. 83*, 7234-7241.

[135] Gorin, G., Martic, P. A., and Doughty, G. (1966) Kinetics of the reaction of N-ethylmaleimide with cysteine and some congeners, *Arch. Biochem. Biophys. 115*, 593-597.

[136] Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions, *Mass Spectrom. Rev. 25*, 663-682.

[137] Mendoza, V. L., and Vachet, R. W. (2009) Probing protein structure by amino acid□ specific covalent labeling and mass spectrometry, *Mass Spectrom. Rev. 28*, 785-815.

[138] Leitner, A., and Lindner, W. (2005) Functional probing of arginine residues in proteins using mass spectrometry and an arginine-specific covalent tagging concept, *Anal. Chem. 77*, 4481-4488.

[139] Zhang, H., Wen, J., Huang, R. Y., Blankenship, R. E., and Gross, M. L. (2012) Mass spectrometry-based carboxyl footprinting of proteins: method evaluation, *Int. J. Mass Spectrom. 312*, 78-86.

[140] Akashi, S., Shirouzu, M., Terada, T., Ito, Y., Yokoyama, S., and Takio, K. (1997) Characterization of the structural difference between active and inactive forms of the Ras protein by chemical modification followed by mass spectrometric peptide mapping, *Anal. Biochem. 248*, 15-25.

[141] Scaloni, A., Monti, M., Acquaviva, R., Tell, G., Damante, G., Formisano, S., and Pucci, P. (1999) Topology of the thyroid transcription factor 1 homeodomain-DNA complex, *Biochemistry 38*, 64-72.

[142] Kelleher, N. L., Lin, H. Y., Valaskovic, G. A., Aaserud, D. J., Fridriksson, E. K., and McLafferty, F. W. (1999) Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry, *J. Am. Chem. Soc. 121*, 806-812.

[143] Xu, P., and Peng, J. (2008) Characterization of polyubiquitin chain structure by middle-down mass spectrometry, *Anal. Chem. 80*, 3438-3444.

[144] Cravello, L., Lascoux, D., and Forest, E. (2003) Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins, *Rapid Commun. Mass Spectrom. 17*, 2387-2393.

[145] Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics, *Nature 422*, 198-207.

[146] Dongré, A. R., Eng, J. K., and Yates, J. R. (1997) Emerging tandem-mass-spectrometry techniques for the rapid identification of proteins, *Trends Biotechnol. 15*, 418-425.

[147] Perry, R. H., Cooks, R. G., and Noll, R. J. (2008) Orbitrap mass spectrometry: instrumentation, ion motion and applications, *Mass Spectrom. Rev. 27*, 661-699.

[148] Tomlinson, A. J., and Naylor, S. (2000) Capillary electrophoresis/mass spectrometry in peptide and protein analysis, *Encyclopedia Anal. Chem.*.

[149] Sharma, S., Zheng, H., Huang, Y. J., Ertekin, A., Hamuro, Y., Rossi, P., Tejero, R., Acton, T. B., Xiao, R., and Jiang, M. (2009) Construct optimization for protein NMR structure analysis using amide hydrogen/deuterium exchange mass spectrometry, *Proteins: Struct., Funct., Bioinf. 76*, 882-894.

[150] Loo, J. A., Berhane, B., Kaddis, C. S., Wooding, K. M., Xie, Y., Kaufman, S. L., and Chernushevich, I. V. (2005) Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex, *J. Am. Soc. Mass. Spectrom. 16*, 998-1008.

[151] Weis, D. D. (2016) *Hydrogen Exchange Mass Spectrometry of Proteins: Fundamentals, Methods, and Applications*, John Wiley & Sons, Hoboken, NJ

[152] Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins, *J. Magn. Reson. (1969) 85*, 393-399.

[153] Hvidt, A., and Nielsen, S. O. (1966) Hydrogen exchange in proteins, *Adv. Prot. Chem. 21*, 287-386.

[154] Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange, *Proteins: Struct., Funct., Bioinf. 17*, 75-86.

[155] Lim, W. K., Rösgen, J., and Englander, S. W. (2009) Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group, *Proc. Natl. Acad. Sci. U.S.A. 106*, 2595-2600.

[156] Glasoe, P. K., and Long, F. A. (1960) Use of glass electrodes to measure acidities in deuterium oxide1,2, *J. Phys. Chem. 64*, 188-190.

[157] Wang, A., Robertson, A. D., and Bolen, D. W. (1995) Effects of a naturally occurring compatible osmolyte on the nternal dynamics of ribonuclease A, *Biochemistry 34*, 15096-15104.

[158] Katta, V., and Chait, B. T. (1993) Hydrogen/deuterium exchange electrospray ionization mass spectrometry: a method for probing protein conformational changes in solution, *J. Am. Chem. Soc. 115*, 6317-6321.

[159] Zhang, Z., and Smith, D. L. (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation, *Protein Sci. 2*, 522-531.

[160] Campbell, S., Rodgers, M., Marzluff, E. M., and Beauchamp, J. (1994) Structural and energetic constraints on gas phase hydrogen/deuterium exchange reactions of protonated peptides with D2O, CD3OD, CD3CO2D, and ND3, *J. Am. Chem. Soc. 116*, 9765-9766.

[161] Pirrone, G. F., Iacob, R. E., and Engen, J. R. (2015) Applications of hydrogen/deuterium exchange MS from 2012 to 2014, *Anal. Chem. 87*, 99-118.

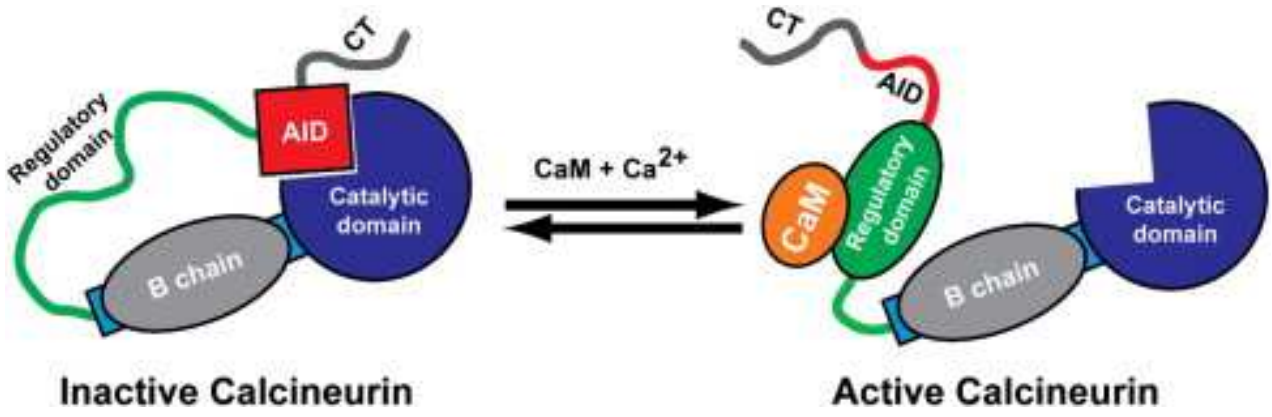
[162] Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: Native-state hydrogen exchange, *Science 269*, 192-197.

[163] Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids, *Quart. Rev. Biophys. 16*, 521-655.

[164] Wang, L., Pan, H., and Smith, D. L. (2002) Hydrogen exchange-mass spectrometry: Optimization of digestion conditions, *Mol. Cell. Proteom. 1*, 132-138.

[165] Hotchko, M., Anand, G. S., Komives, E. A., and Ten Eyck, L. F. (2006) Automated extraction of backbone deuteration levels from amide H/2H mass spectrometry experiments, *Protein Sci. 15*, 583-601.

Chapter 2: Structural basis for activation of calcineurin by calmodulin†



† Reprinted from Journal of Molecular Biology, 415/2, Rumi-Masante J., Rusinga F.I., Lester T.E., Dunlap T.B., Williams T.D., Dunker A.K., Weis D.D., Creamer T.P., Structural basis for activation of calcineurin by calmodulin, 307-317, Copyright (2011), with permission from Elsevier.

2.1 Introduction

Calcineurin (CaN) is a serine/threonine phosphatase originally identified by Wang and Desai[1], Watterson and Vanaman[2], and Klee and Krinks.[3] In response to elevated calcium levels, calmodulin (CaM) binds to CaN leading to its activation. CaN plays essential roles in T cell activation, nervous system development and function, and cardiac growth.[4] As a result, dysregulation of CaN has been implicated in a number of disease states including cardiac hypertrophy[5,6], Alzheimer's disease[7] and Down syndrome.[8] CaN is also the target for the immunosuppressant drugs FK506 and cyclosporin A.[9] Despite its importance, remarkably little is known regarding the mechanism of CaN's activation by CaM.

CaN is a heterodimer consisting of an ~60kDa A chain and a 19kDa B chain.[10] There are three isoforms of the CaN A chain: the α isoform which is the dominant form in neurons; the broadly-distributed β isoform; and the testis-specific γ isoform.[11] The β and γ isoforms differ from the α isoform primarily through N- and C-terminal extensions respectively. The CaN A chain consists of a catalytic domain, B chain binding domain, a 95 residue regulatory domain (RD) including the CaM binding region near its N-terminal end, an autoinhibitory domain (AID), and a short C-terminal domain (CT; Figure 2.1a). There are two isoforms of the B chain, with isoform 1 being associated with the αA and βA chains. The B chain is homologous to CaM and is known to bind four calcium ions.[12]

At low calcium concentrations CaN exists in an inactive state, with its AID bound in the active site cleft. Upon an increase in calcium concentration, CaM binds four calcium ions, resulting in binding to the CaN RD. CaM binding causes release of the AID and activation of the phosphatase (Figure 2.1b).[13] This mechanism of activation is similar to that observed for the

calmodulin-dependent protein kinases, although in the kinases the calmodulin binding site is immediately adjacent in sequence to the autoinhibitory domain.[14]

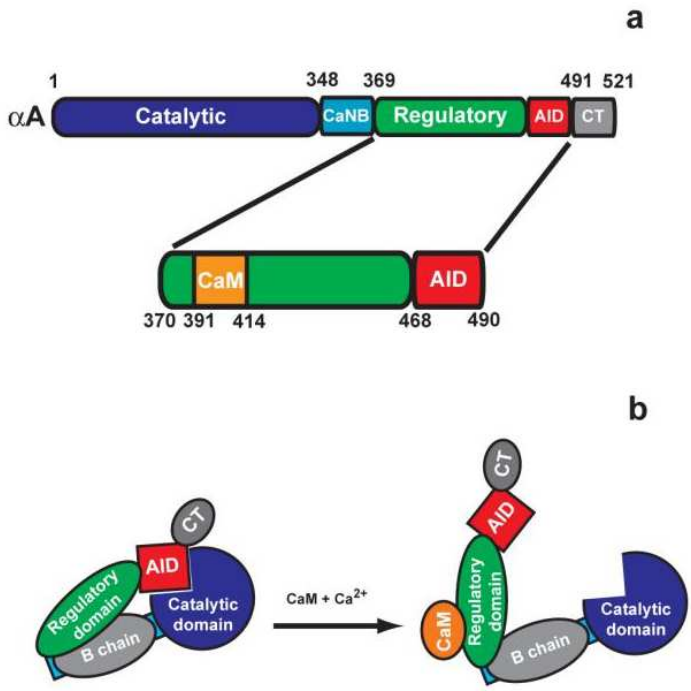


Figure 2.1 Structure and activation of human αCaN. a. Domain structure of the αCaN A chain highlighting the regulatory domain, with CaM binding region, and autoinhibitory domain. b. Model for CaN activation by CaM.

Published data suggest that CaN's RD is at least partially disordered in the absence of CaM. Manalan and Klee[11] showed that trypsin readily hydrolyzed the RD, implying that it was readily accessible. Further hydrolysis experiments by Yang and Klee[13] confirmed the flexible and accessible nature of the RD. In addition, in the crystal structure of full-length αCaN the AID is seen bound in the active site of the catalytic domain, but electron density for the remainder of the RD, and for the CT, was missing.[15] Following up on these observations, Dunker and co-workers noted the potentially disordered nature of the RD in 1997.[16]

Recently Shen et al.[17] generated a fragment of αCaN corresponding to the RD, AID and CT that was shown, via Fourier transform infrared spectroscopy (FTIR), to be largely

unstructured. Using acrylamide collisional quenching of fluorescence from the four intrinsic tryptophan residues in αCaN, these authors also generated data supporting the earlier findings of Yang and Klee[13] that the CaM binding region of CaN interacts with its B chain binding domain in the absence of calcium and CaM. Shen et al.'s fluorescence experiments also indicated that the CaM-bound RD does not interact with the remainder of αCaN.[17]

The limited tryptic digestion experiments of Manalan and Klee[11] suggested that the RD of CaN undergoes a large conformational change upon CaM binding. They found that the RD was protected from trypsin hydrolysis when αCaN was pre-incubated with CaM. What is remarkable about this observation is that the CaM binding region in the RD spans just 24 residues, while the RD is 95 residues in length (Figure 2.2).

Figure 2.2 Sequence of the αCaN RD-AID-CT construct used in this work. Residues in gray were added to aid with expression and purification. The RD is denoted by green residues, with the CaM binding region highlighted in orange. The AID sequence is shown in red and the CT in black. Residue numbering is based on that for the full-length αCaNA chain.

Fifty-two residues separate the CaM binding region from the start of the AID, including four potential trypsin cleavage sites. The CaM-induced conformational change likely then

involves more of the RD than just the CaM binding region, a hypothesis that is consistent with previous biochemical and biophysical work.[11,13,17,18]

In this work we present data from fluorescence experiments demonstrating that the RD does not interact with the remainder of the human αCaN heterodimer when CaM is bound. Using circular dichroism (CD) spectroscopy and hydrogen/deuterium-exchange mass spectrometry (HXMS) we show that a fragment corresponding to the αCaN RD through C-terminus (RD-AID-CT) is completely disordered in solution. With the exception of the AID, this region is also likely to be disordered in full length αCaN in the absence of CaM when the B chain is fully loaded with calcium. Using CD we show that the RD-AID-CT fragment gains substantial α-helix content upon CaM binding. Employing HXMS we demonstrate that the conformational change that occurs upon CaM binding is restricted to the RD and that the AID and CT are not involved. Furthermore, these data identify two regions of the RD that become structured upon the binding of CaM. The most highly protected from exchange is the CaM binding region. The second, less well-protected region is located within an ~35 residue stretch C-terminal to the CaM binding region. These data indicate that CaM induces folding of the RD on a scale that has not previously been observed for CaM binding to a target protein.

2.1.1 Statement of collaboration

This chapter describes a collaborative study between the Weis lab and Dr. Trevor Creamer's research group, from the University of Kentucky. The resulting publication is reproduced here with permission from the corresponding author, Dr. Creamer and under license from the publisher, Elsevier. Limited proteolysis, fluorescence and CD measurements of calcineurin constructs were performed at the University of Kentucky. Intact protein HX-MS measurements

on CaN-RD-AID-CT were performed by David Weis. Peptide HX-MS measurements were performed on CaN-RD-AID-CT, by the author of this dissertation, Farai Rusinga, of the Weis research group. In this chapter, HX-MS data is presented together with the other biophysical measurements to provide the complete context under which a mechanism for calcineurin activation is proposed. In summary, limited proteolysis and fluorescence data suggested that the CaN-RD-AID-CT adopts structure when bound to calmodulin in the presence of calcium. CD data showed that the proportion of CaN-RD-AID-CT residues in α-helices increases upon binding calmodulin, and further provided evidence that helicity extends beyond the calmodulin binding domain. HX-MS data was crucial in identifying the sites that adopt helicity, which are the calmodulin-binding domain and an adjacent region of CaN-RD-AID-CT. Furthermore, rapid HX at all residues of free CaN-RD-AID-CT confirmed that this region of calcineurin is a random coil in the absence of calmodulin. HX-MS measurements were then used to propose the mechanism of calcineurin activation, the ultimate goal of the study presented in this chapter.

2.2 Experimental

2.2.1 Proteins, peptide and buffer

The plasmid pETagHisCN, containing the human αCaNA (with N-terminal His_6 tag) and B1 genes, together forming αCaN, was obtained from Addgene (Cambridge, MA). This was transformed into E. coli BL21 (DE3) CodonPlus RIL cells (Agilent Technologies, La Jolla, CA) for expression. αCaN was purified on a Ni-NTA column followed by a CaM-sepharose column (GE Healthcare, Piscataway, NJ). αCaN mutants were generated using the Stratagene QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies, La Jolla, CA). An E. coli-codon optimized gene for the human sequence RD-AID-CT construct was synthesized by

Genscript (Piscataway, NJ). This was cloned into the pET303/CT-His vector which adds a C-terminal His_6 tag (Invitrogen, Carlsbad, CA) for expression in E. coli BL21 (DE3). After expression, the RD-AID-CT was purified on a Ni-NTA column followed by a CaM-sepharose column. Human CaM was expressed from the pETCaMI vector and purified on a 2-trifluoromethyl-10-aminopropyl phenothiazine-sepharose (TAPP-sepharose) column.[35] The TAPP-sepharose was synthesized at the University of Kentucky, Center for Structural Biology Chemistry Core Facility (supported in part by funds from NIH National Center for Research Resources (NCRR) grant P20 RR020171). All expressed proteins had their identities confirmed via mass spectrometry. Protein concentrations were determined using the bicinchoninic acid (BCA) assay.[36]

The pCaN peptide (sequence ARKEVIRNKIRAIGKMARVFSVLR) corresponding to the CaM binding region in the RD of αCaN, was purchased from Genscript (Piscataway, NJ). This was purified using reverse-phase HPLC and its identity confirmed using mass spectrometry.

All experiments were conducted using a buffer consisting of 20mM Tris, 200mM NaCl, 2mM $CaCl_2$ at a pH of 7.5. All reagents used for this buffer were obtained from Sigma (St. Louis, MO) and were of the highest purity.

2.2.2 Limited proteolytic digests

Tryptic digests were performed by adding sequencing-grade trypsin (Promega, WI) to protein solutions. A ratio of 1:200 trypsin to protein was utilized. Aliquots were removed at various time-points and hydrolysis quenched by the addition of excess soybean trypsin inhibitor (Sigma-Aldrich, St. Louis, MO). Resulting digests were run out on SDS-PAGE gels and visualized using Coomassie blue staining. Bands on the gels were excised and analyzed using

MALDI-TOF mass spectrometry at the University of Kentucky Center for Structural Biology Proteomics Core.

2.2.3 Fluorescence

Fluorescence spectra were collected on a Perkin-Elmer LS55 steady-state fluorimeter using 1cm pathlength cuvettes. Proteins were at a concentration of 1μM for all tryptophan fluorescence experiments. Samples were excited at 290nm and spectra were collected from 305nm to 430nm with excitation and emission bandwidths of 10nm.

For fluorescence anisotropy experiments Cy3 was coupled to the RD-AID-CT construct via the His_6-tag on the protein. To do this, Cy3 was covalently linked to nickel(II)-nitrilotriacetic acid (Ni^{2+}-NTA) to form $(Ni^{2+}\text{-NTA})_2$-Cy3 using the protocol of Zhao et al.[20] Cy3 was obtained from GE Healthcare (Piscataway, NJ) and all other required reagents, including Nα,Nα-bis(carboxymethyl)-L-lysine hydrate for (Ni^{2+}-NTA) synthesis, from Sigma (St. Louis, MO). $(Ni^{2+}\text{-NTA})_2$-Cy3 was coupled to RD-AID-CT by mixing a two-fold of excess of the fluorophore with protein and incubating at 4°C with shaking for three hours. Uncoupled $(Ni^{2+}\text{-NTA})_2$-Cy3 was separated from labeled protein, RD-AID-CT-fl, on a 10ml, 1cm diameter G-10 Sephadex column. Anisotropy measurements were made using a 1cm cuvette and samples with a Cy3 concentration of 0.5 μM (total RD-AID-CT concentration of 1.5 μM) and a two-fold excess of CaN373stop and CaM over RD-AID-CT-fl when added. Excitation was performed at a wavelength of 550nm and emission monitored at 570nm. Excitation and emission bandwidths were both set to 5nm.

2.2.4 Circular dichroism

Circular dichroism (CD) spectra were collected using a Jasco J-810 spectropolarimeter equipped with a Peltier heating block. Samples containing 10μM protein were placed in a 1mm pathlength cuvette, with reported spectra being the average of four scans at a scan speed of 50nm/min and a temperature of 20°C. Errors are estimated to be no more than 3%. Secondary structure content was estimated from the CD spectra using the CONTIN/LL deconvolution program.[21,22] The thermal melt of the RD-AID-CT construct was conducted with a heating rate of 1°C min^{-1} and was monitored at a wavelength of 222nm.

2.2.5 Hydrogen/deuterium-exchange mass spectrometry (HXMS)

2.5.5.1 Materials

Buffers were prepared using tris, HCl, NaCl, NaOH obtained from Fisher Scientific (Hanover Park, IL), and $CaCl_2$ from Sigma (St Louis, MO). HPLC solvents were prepared using Optima LC/MS grade acetonitrile and water from Fisher Scientific (Fair Lawn, NJ) and 99+% formic acid from Thermo Scientific (Rockford, IL).

2.5.5.2 Preparation of RD-AID-CT Peptic Peptides

RD-AID-CT peptic peptides were prepared by diluting 200μL RD-AID-CT stock (20.3μM) to 500μL in 0.1% formic acid. The solution was passed through the immobilized pepsin column in a 4°C refrigerator using a syringe pump at 50μL min^{-1} The first 100μL of eluent was discarded, and then the rest was collected along with a 150μL 0.1% formic acid after-wash. 100μL aliquots of digested peptide solution were frozen on liquid nitrogen and stored at –

80°C. These peptides were used to prepare totally deuterated RD-AID-CT and to confirm peptide assignments.

2.5.5.3 Hydrogen/Deuterium Exchange

For intact hydrogen-deuterium exchange, 6.7µL aliquots containing RD-AID-CT at 9µM and CaM at 18µM (or an equivalent volume of buffer) were held at 4°C for one hour. Following incubation, individual aliquots were diluted 20-fold with D_2O buffer (20mM Tris/200mM NaCl/10mM $CaCl_2$, pD 7.51). Exchange proceeded for between 5 sec and 2 hours at 4°C in a thermostated block (IC20 Ecotherm Heating/Chilling Plate, Torrey Pines Scientific, La Jolla, CA). Totally deuterated RD-AID-CT was prepared by diluting the sample 20-fold with 4M GdnDCl, exchanged for two hours at room temperature. 4 M GdnDl was prepared by several rounds of dissolution/lyophilization of GdnHCl in D_2O. The exchange reactions were quenched by 1:1 dilution with 200mM 4 °C sodium phosphate buffer (pH 2.6). Samples were immediately flash-frozen on liquid nitrogen and held at –80°C.

For peptide-level hydrogen-deuterium exchange, RD-AID-CT (9µM) was incubated with a two-fold molar excess of CaM (18µM) or an equal volume of H_2O buffer (20mM Tris/200mM NaCl/10mM $CaCl_2$, pH 7.51) for at least one hour at 25°C. Aliquots (5.7µL, containing 50pmol of RD-AID-CT) were diluted 20-fold with D_2O buffer (20mM Tris/200mM NaCl/10 mM $CaCl_2$, pD 7.51). Samples were labeled for between 5 seconds and 24 hours at 25°C. After labeling, the samples were quenched to a pH of 2.4 by addition of 0.1M HCl (6.8µL), frozen on liquid nitrogen, and stored on dry ice at –80°C. Undeuterated samples were prepared similarly and diluted 20-fold with H_2O buffer instead of D_2O buffer. Totally deuterated samples were prepared

from RD-AID-CT peptic peptides, deuterated and quenched, then stored as described above. Peptide-level measurements were completed in duplicate.

2.5.5.4 LC/MS Analysis

All LC/MS analysis was carried out using a custom refrigerated LC system described previously.[23] Individual samples were thawed by hand immediately prior to loading onto the LC system.

Intact deuterated samples were desalted on a reversed-phase trap (Jupiter C4, Phenomenex, Torrance, CA, self-packed 1×10 mm cartridge) for 3 min using 200µL min^{-1} of 0.1% formic acid followed by elution using the following gradient: 20% B (0-3 min), 30% B (3.5 min), 45% B (6 min), 95% B (7 min), 95% B (8 min), 20% B (9 min). Here A was 0.1% formic acid and B was acetonitrile/water/formic acid (90%/10%/0.1%).

Peptide level samples, were digested online at 200µL min^{-1} with 0.1% formic acid as the carrier using pepsin (from porcine gastric mucosa, Sigma, St. Louis, MO) immobilized on POROS 20AL (Applied Biosystems, Carlsbad, CA)[37,38] packed in a 2.1×50 mm stainless steel column. The resulting peptides were trapped and desalted on a reversed-phase trap (1×10 mm self-packed with Jupiter Proteo C12 media obtained from Phenomenex, Torrance, CA) over 4 minutes. The peptides were eluted from the trap and separated using a Zorbax 300SB-C18, 1×50 mm column (Agilent Technologies, Santa Clara, CA). The elution gradient, with solvents A (0.1% formic acid) and B (90:10:0.1% acetonitrile:water:formic acid), was programed as follows: 4 min at 10% B, 10-20% B over 30 secs, 20-40% B over 5 min, 40% B for 30 secs, 40-95% B over 1.5 min, hold at 95% B for 1 min, then brought back down to 10% B over 30 s.

Masses of deuterated samples were measured using a time of flight mass analyzer with an electrospray ionization source (model 6220, Agilent Technologies, Santa Clara, CA). All mass spectra were collected in positive ESI mode, in the 2GHz Extended dynamic range with a capillary voltage of 4000V, drying gas flow of 10 L min^{-1} and temperature of 325°C. The fragmentor was set to 200V for intact protein measurements and 150V for peptides. Mass spectra were acquired using Agilent MassHunter Acquisition and analyzed using Agilent MassHunter Qualitative Analysis software (Versions B.03.01 & B.04.00). The extent of deuteration (as a percentage of the exchangeable amides) was calculated with back-exchange corrections as previously described by Zhang and Smith.[39] Here, the number of exchangeable amides accounts for all non-proline residues from the third residue to the C-terminus of the peptide. In cases where totally deuterated peptides were not detected (see Figure 2.A), the average deuterium uptake by free RD-AID-CT was used instead of the totally deuterated sample.

2.5.5.5 Confirmation of Peptide Identities

RD-AID-CT peptic peptides were injected onto the C12 trapping column by isocratic flow as described above (without passing through a pepsin column). For optimum chromatographic separation, a long gradient run was set up as follows: 10% B for 4 min, 10-45% B over 41 min, 45% B for 5 min, ramp up 45-95% B over 5 min, 95% B for 5 min, down to 10% B over 2 min. Identities were confirmed by accurate mass measurements using internal mass correction in 4GHz high resolution mode. Peptides that could not be definitively assigned on the basis of accurate mass alone (within a 10 ppm mass window) were assigned on the basis of their CID fragmentation using MS^E and MS/MS. Identity confirmation experiments were performed on a Synapt G2 QTOF MS with a NanoAcquity UPLC (Waters, Millford, MA) with an ESI source, at the Structural Biology Center (University of Kansas). Peptides were separated through

a Zorbax 300SB-C18, 0.3×50 mm, 3.5µm, 300Å analytical column (Agilent Technologies, Santa Clara, CA) at a flow rate of 10µL min^{-1} The gradient, for solvents A (99:1:0.08% water:acetonitrile:formic acid) and B (80:10:10:0.06% acetonitrile:isopropanol:water:formic acid), was set up as follows: 1-10% B over 1 min, 10-30% B over 34 min, ramped up to 80% B in 1 min and returned to 1.0% B in 1 min. Fragmentation was performed in MS^E mode using low (30V) and high (50V) collision energies and in data-dependent acquisition mode using mass/charge-dependent collision energies. In MS/MS mode, individually-optimized collision energies were used. Mass spectra were analyzed for expected a/b/y ion fragments, corresponding neutral losses and immonium ions using MassLynx (V4.1) software. Ambiguous (isobaric) peptide assignments are listed in Table 2.B.

2.3 Results

The initial evidence for the αCaN RD being disordered and undergoing a conformational change upon CaM binding was a limited tryptic digestion performed by Manalan and Klee.[11] We repeated this experiment (Figure 2.C). In the absence of CaM, the αCaNA chain is partially hydrolyzed, resulting in the appearance of an ~45kD fragment within 5 minutes. The CaNB1 chain is not hydrolyzed by trypsin in the same timeframe. If αCaN is preincubated with an excess of CaM, it is largely protected from digestion (Figure 2.C).

A PONDR disorder prediction[19] for αCaNA indicates that the RD-AID-CT region is disordered, with the exception of a short region at the junction between the RD and AID (Figure 2.3). A portion of the catalytic domain is also predicted to be disordered. This region is well-ordered in the αCaN crystal structure, in part packing against the B chain[15], a situation PONDR cannot take into account.

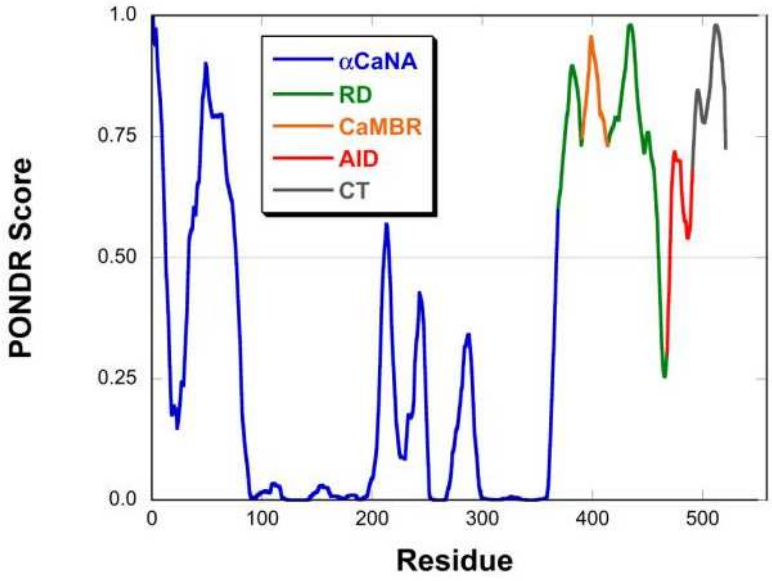


Figure 2.3 PONDR prediction[19] of αCaNA. The catalytic domain is blue, RD green, CaM binding region orange, AID red and CT dark gray.

2.3.1 The CaM-bound RD, AID and CT do not interact with the remainder of αCaN

In order to study interactions between CaM and an isolated fragment consisting of the RD, AID, and C-terminal domain (RD-AID-CT; Figure 2.2) we first needed to determine to what extent the CaM-bound αCaN RD interacts with the remainder of the phosphatase. Potential interactions were probed in two ways: fluorescence of tryptophans introduced into solvent-exposed positions on both the A and B chains of full-length αCaN and fluorescence anisotropy of a dye-labeled RD-AID-CT construct interacting with a truncated αCaN.

For the tryptophan fluorescence experiments, solvent-exposed phenylalanine and tyrosine residues in αCaN were identified and mutated individually to tryptophan. This resulted in four mutations: A:Y341W and A:F356W in the A chain, and B:F72W and B:Y106W in the B chain. Note that αCaN possesses four endogenous tryptophans. Fluorescence spectra for wild-type

αCaN and each of the four mutants were collected both in the absence and presence of saturating levels of CaM (Figure 2.4).

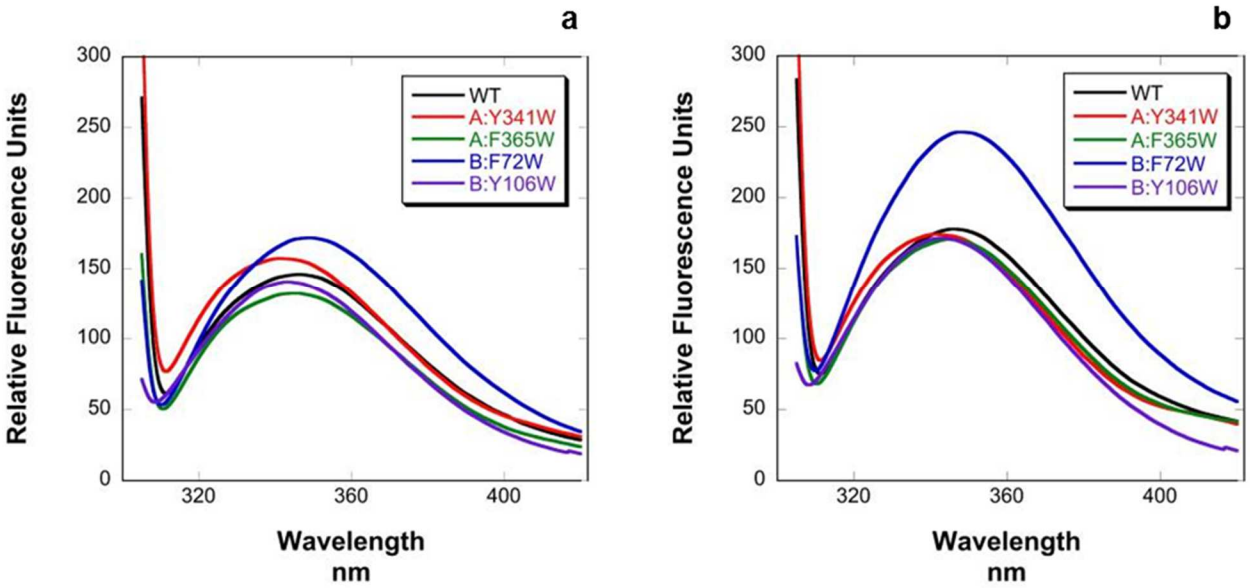


Figure 2.4 Fluorescence emission spectra for wild type αCaN and four mutants with introduced tryptophan residues in the a. absence and b. presence of CaM

Although differences in fluorescence intensity were observed, the wavelength of maximal emission did not change upon CaM binding for wild-type or any of the four mutants (Table 2.1) indicating that none of the introduced, or endogenous, tryptophans experienced changes in solvent accessibility upon complex formation.

Table 2.1 Fluorescence maximum emission wavelengths (nm) in the absence and presence of CaM for wild type αCaN and the four mutants in which a single tryptophan is introduced into a solvent exposed position.

	-CaM	+CaM
WT	346.5	346
A:Y341W	341.5	341
A:F356W	345	345
B:F72W	349	348
B:Y106W	343	343

A truncated αCaN (αCaN373stop) was generated by introducing a stop codon immediately following the codon for residue 373. Residue 373 immediately follows the B chain binding domain (Figure 2.1a) allowing αCaN373stop to bind the B chain. The presence of the B chain, after purification of αCaN373stop, was confirmed by visualization on SDS-PAGE. Following the protocol of Zhao et al.[20], the RD-AID-CT construct was incubated with $(Ni^{2+}\text{-NTA})_2$-Cy3 to generate a fluorescently labeled construct, RD-AID-CT-fl, for use in fluorescence anisotropy experiments. In the absence of CaM, RD-AID-CT-fl binds to αCaN373stop as indicated by an increase in anisotropy (Figure 2.5). When an excess of CaM is present, the anisotropy decreases to a similar level as that obtained for the CaM-bound RD-AID-CT-fl, suggesting the interaction between the RD-AID-CT-fl and αCaN373stop has been abolished. Notably CaM-bound RD-AID-CT-fl has a similar anisotropy as the isolated RD-AID-CT-fl (Figure 2.5).

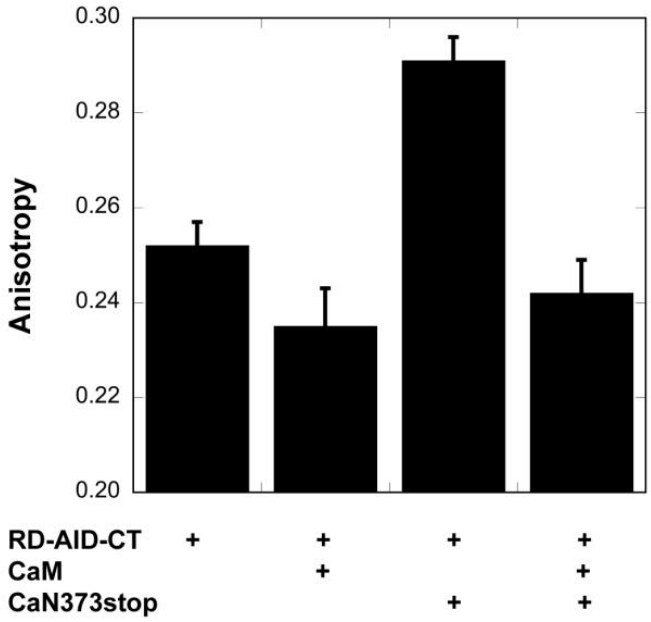


Figure 2.5 Fluorescence anisotropy results for a fluorescently labeled RD-AID-CT construct in the presence of a two-fold excess CaM, the truncated αCaN373stop, and both.

2.3.2 The RD-AID-CT undergoes a disorder to order transition upon CaM binding

The secondary structure content of the RD-AID-CT construct in the absence of CaM was examined via CD spectroscopy (Figure 2.6a). The spectrum obtained is characteristic of an unstructured polypeptide chain. CONTIN/LL analysis[21,22] of the spectrum suggested the presence of predominantly unstructured chain (Table 2.2). Heating the sample did not result in an observable unfolding (Figure 2.6a inset), consistent with the interpretation of an unstructured chain. HXMS data for the RD-AID-CT construct in the absence of CaM reinforced this observation (Figure 2.6b). When corrected for back-exchange these data indicate that essentially all backbone amide protons exchanged for deuterons within five seconds. HXMS data collected in the presence of 4M GdnDCl lend further support to the assertion that the RD-AID-CT is devoid of stable structure (Figure 2.6b).

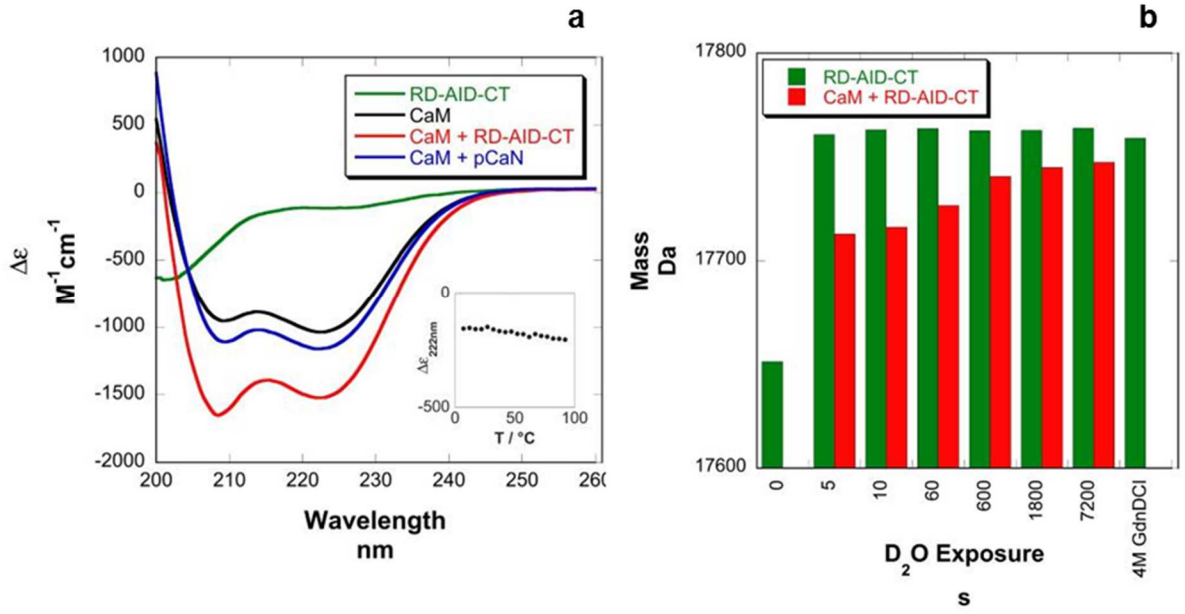


Figure 2.6 Experimental evidence for the disordered nature of the RD-AID-CT and its acquisition of α-helical structure upon CaM binding. a. CD spectra for the RD-AID-CT, CaM, the CaM:RD-AID-CT complex, and a complex of CaM bound to pCaN (the CaM binding domain from αCaN). b. HXMS data for the intact RD-AID-CT in the absence and presence of CaM. The data shown are from single experiments and are not corrected for back-exchange

Table 2.2 CONTIN/LL 21 analyses of CD spectra for the RD-AID-CT, CaM and CaM:RD-AID-CT complex. Results are expressed as the average number of residues (<Naverage>) in each secondary structure type and are rounded to the nearest whole number.

<Naverage>	RD-AID-CT	CaM	CaM + RD-AID-CT	Net change
α-helix	12	103	165	50
β-strand	57	1	29	-29
turns	33	16	47	-2
coil	63	29	72	-20

2.3.3 CaM-mediated ordering is localized to the RD

Trypsin digestion of the RD-AID-CT preincubated with CaM resulted in two fragments of ~13 kDa mass (Figure 2.7). MALDI-TOF mass spectrometric analysis of these two fragments revealed that both encompassed the entire RD, with one cleaved after the lysine that denotes the end of the RD (K466), and the second having been cleaved after the first lysine within the AID (K474, see Figure 2.2). We do not observe cleavage following any of the other nine basic residues in the RD. These data indicate that the AID and CT are both readily accessible to trypsin and are likely disordered, but that the RD is protected from hydrolysis in the CaM:RD-AID-CT complex.

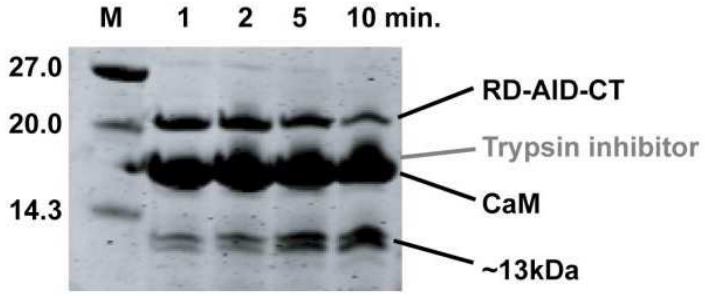


Figure 2.7 SDS-PAGE gel showing the results of a tryptic digest of the RD-AID-CT construct in the presence of a ten-fold excess of CaM.

The CD spectrum of an equimolar mixture of RD-AID-CT and CaM is indicative of a large increase in α-helix content relative to CaM alone (Figure 2.6a). An equimolar mixture of CaM and pCaN, a 24 residue peptide corresponding to the CaM-binding region of αCaN, also indicates an increase in α-helix content, but to a lesser extent than the RD-AID-CT:CaM complex (Figure 2.6a). CONTIN/LL analysis[21,22] of the CaM:RD-AID-CT spectrum suggests that ~50 residues in the RD-AID-CT become α-helical upon CaM binding (Table 2). HXMS data collected for the full-length RD-AID-CT preincubated with a two-fold excess of CaM are shown in Figure 2.6b. Here, the level of back-exchange (~30%) can be estimated using deuterium uptake by free RD-AID-CT as a totally deuterated sample. At short exchange times, after correction for back exchange, ~50 backbone amides are protected, in good agreement with the CD data. After two hours of exchange approximately 20 residues remain protected. These results indicate that CaM binding induces modest protection in approximately 30 residues and much stronger protection in approximately 20 residues.

2.3.4 Two regions of the RD fold upon CaM binding

To localize secondary structure, the deuterium-labeled RD-AID-CT construct was digested using pepsin after quenching the hydrogen-deuterium exchange reaction. Deuterium incorporation in the resulting peptides was measured by mass spectrometry following a short HPLC separation under quenched conditions.[23] Patterns of exchange are shown in a HXMS heat map (Figure 2.8). Deuterium uptake curves for the individual peptide fragments are shown in Figure 2.A. All backbone amides in the RD-AID-CT construct in the absence of CaM exchange within five seconds. When preincubated with a two-fold excess of CaM, the RD-AID-CT exchange patterns are quite different. The CaM-binding region is very highly protected, with significant protection levels even after 24 hours of exposure to D_2O. One other region of

protection is observed: an ~35 residue region C-terminal to the CaM-binding region. Note that not all residues in this region are protected from exchange - the resolution is limited by the lack of overlapping pepsin fragments generated in this region. Both regions of protection are within the regulatory domain (RD), in very good agreement with MALDI-TOF analysis of the limited trypsin digestion data (Figure 2.7). Combining the CD, HXMS and proteolysis data, it would appear that a large fraction of the RD becomes ordered, largely α-helical, upon CaM binding.

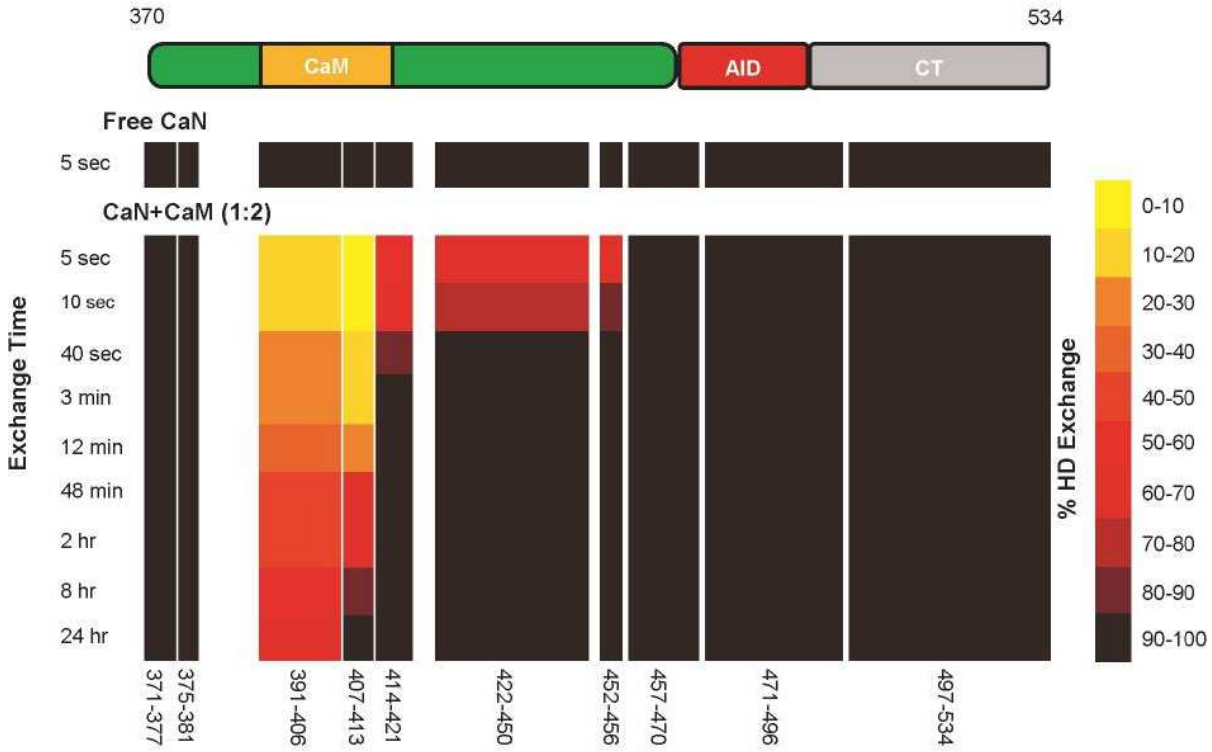


Figure 2.8 HXMS heat map for the RD-AID-CT in the absence and presence of a two-fold excess of CaM. Structure of the RD-AID-CT construct is shown at top for reference. The map was constructed using a minimal set of ten non-overlapping peptides for which complete duplicate data were available. See Figure 2.A for complete deuterium uptake kinetics for all RD-AID-CT peptides.

2.4 Discussion

The limited tryptic digestion data of Manalan and Klee[11] suggested that the RD of αCaN was at least partially disordered in the absence of CaM, and that it became ordered upon CaM binding. Hubbard and Klee[24] used limited clostripain digestion to map out the domain structure

of the αCaN A chain, this data also hinted at the disordered nature of the RD. Our limited tryptic digest data on full-length αCaN (Figure 2.B) is in good agreement with these earlier studies. The first crystal structures of full-length CaN were solved by Kissinger et al.[15] In these structures, the electron density for the RD and CT was missing, indicating these domains were mobile in the crystals and again hinting at disorder. Further evidence for disorder comes from a PONDR prediction[19] for the αCaN A chain (Figure 2.3). Our CD and HXMS data for the isolated RD-AID-CT construct (Figure 2.6) support the hypothesis that the RD is disordered in the absence of CaM. All backbone amide protons in the RD-AID-CT construct exchange with deuterons within five seconds indicating a lack of detectable stable secondary structure. The FTIR data of Shen et al.[17] for a similar fragment of αCaN also indicates that this region is disordered in the absence of CaM.

It was important to determine whether the binding of CaM to the RD-AID-CT construct resulted in interactions between this complex and the remainder of αCaN. Shen et al.[17] measured acrylamide quenching of the fluorescence of αCaN's four intrinsic tryptophans. They concluded that the solvent accessibility is unchanged for the tryptophans upon CaM binding. In this work we took two approaches. In the first approach, we substituted four solvent-exposed aromatic residues with tryptophan and determined the effects of CaM binding upon their fluorescence emission. Although some changes in emission intensity were observed (Figure 2.4), most notably for the B:F72W mutant, the wavelength of maximum emission did not change for any of these mutants upon the binding of CaM (Table 1). These data indicate that none of the introduced tryptophans are buried when CaM binds to the RD.

In the second approach, fluorescently labeled RD-AID-CT was used in anisotropy measurements. In the absence of CaM, RD-AID-CT-fl binds to the truncated αCaN373stop

(Figure 2.5). In the presence of an excess of CaM, the observed anisotropy is reduced to that of the CaM:RD-AID-CT-fl complex. These data indicate that binding of RD-AID-CT-fl to αCaN373stop is abolished when CaM binds to the RD. We hypothesize that the observed binding of RD-AID-CT-fl to αCaN373stop in the absence of CaM is due to the AID binding within the active site of αCaN373stop's active site. Soderling and co-workers[25,26] have shown that a peptide corresponding to the AID can bind to and inhibit CaN, lending support to this hypothesis. All of our experiments were performed in the presence of an excess of calcium, precluding interactions between the CaM binding region and the B chain binding domain.[13]

It is notable that the anisotropy measured for the CaM:RD-AID-CT-fl complex is very similar to that for the RD-AID-CT-fl construct alone (Figure 2.5). Anisotropy is a function of both the rotational diffusion of the molecule or complex, and the flexibility of the chain to which the fluorophore is attached.[27] The $(Ni^{2+}\text{-NTA})_2$-Cy3 fluorophore is bound to the His_6-tag located at the C-terminus of RD-AID-CT (Figure 2.2). Our CD and HXMS experiments (Figures 2.6 and 8) indicate that when CaM binds to RD-AID-CT, the AID and CT remain disordered. Therefore, the $(Ni^{2+}\text{-NTA})_2$-Cy3 fluorophore is attached to a long stretch of disordered chain whose flexibility is unaffected by CaM binding, providing an explanation for why no significant change in anisotropy is observed. In contrast, when the RD-AID-CT-fl construct interacts with CaN373stop, it likely does so via binding of the AID into the CaN active site cleft[18], leading to structuring of the AID[15] and subsequent reduction in flexibility of the chain to which the fluorophore is linked, leading to an increase in anisotropy.

Notably, when CaM binds to the RD-AID-CT construct there is a conformational change that precludes binding to αCaN373stop via the AID. This conformational change appears to be a large scale folding of the RD, with subsequent gain in α-helix content (Figure 2.6a and Table

2.2). Trypsin digestion and HXMS data suggest that the AID and CT are disordered (Figures 2.7 and 2.8). Perrino[18] demonstrated that a construct similar to our RD-AID-CT inhibits CaN activity and that inhibition is relieved by addition of CaM. A model that explains all of these data would have the C-terminal region of the RD folded in the CaM:RD complex, preventing interactions between the AID and the active site via steric hindrance.

There is remarkably good agreement between the CD estimate of ~50 residues becoming ordered (Figure 2.6a and Table 2.2) and early time points in the HXMS on the intact CaM:RD-AID-CT complex (Figure 2.6b). From the heat map (Figure 2.8) and peptide exchange uptake curves (Figure 2.A) generated after pepsin digestion of the exchanged RD-AID-CT it is clear there are two regions of protection within the RD. The CaM binding region is very well-protected from exchange, as might be expected given CaM binds this region with a K_d in the picomolar range.[28,29] The second region of protection is C-terminal to the CaM binding region (Figure 2.8). This region is less well-protected, explaining the decrease in protection with time illustrated in Figure 2.6b. Later time points in the HXMS data on the intact CaM:RD-AID-CT complex (Figure 2.6b and Figure 2.A) could be interpreted as indicating this latter protected region has a significant fraction of its ensemble unstructured. This would be at odds with the CD data (Figure 2.6a and Table 2.2). If the region of lower protection were unfolded a significant fraction of time, ensemble-averaged CD data should give lower estimates of the number of residues in the α-helical conformation. However, as noted by Henkels and Oas[30] and Keppel et al.[31], the kinetics of hydrogen-deuterium exchange can be dominated by unprotected states with small Boltzmann weights when the inter-conversion between protected and unprotected conformers is rapid relative to the hydrogen-deuterium exchange time-scale. If this is the case for

the protected region C-terminal to the CaM binding region, then there is no disagreement between the CD and HXMS data.

Perrino[18] has shown that the thirty-eight residue region of the RD immediately N-terminal to the AID plays a role in the autoinhibition of CaN. Soderling and co-workers[25] had shown earlier that deletion of the AID, but not the RD, leads to only partial activation of CaN. Full activation is achieved in the absence of CaM only when both the AID and the thirty-eight residues preceding it were deleted. This suggests there is an interaction between the chain immediately preceding the AID and the remainder of CaN, but this is likely not a stable interaction given the ease with which the RD can be proteolytically digested (Figure 2.B).[11,24] This second autoinhibitory region overlaps with the region of lower protection observed in the HXMS data for the CaM-bound RD-AID-CT (Figure 2.8). The region of limited exchange protection appears to be part of the large scale folding of the CaM-bound RD, further explaining how this folding leads to CaN activation.

Our data lead us to propose the following model for αCaN activation by CaM. In the absence of CaM, the AID is ordered and bound within the active site cleft of αCaN, while the RD and CT are largely disordered. In the absence of calcium, the CaM binding region[13] and putative second autoinhibitory region[18] may be in contact with the B chain and its binding domain. Upon an increase in calcium concentration and subsequent CaM binding, the RD folds, gaining significant α-helical structure in the CaM binding and second autoinhibitory regions. This folding event provides the driving force that pulls the AID from the active site. In the CaM-bound state the released AID and CT are largely disordered.

Coupled binding and folding, such as that described by our model, is thought to be a common feature of disordered regions in proteins.[32,33] This is the mechanism by which CaM appears to activate calmodulin-dependent protein kinases[14], suggesting that coupled folding and binding might be a general mechanism by which CaM activates its target enzymes. Dunker and colleagues have recently shown that CaM binding sites are typically within disordered regions in proteins[34], lending support to this idea. The major difference between the kinases and CaN is that the CaM binding sites and AID's in the kinases are immediately adjacent in sequence, whereas there are 52 residues between the CaM binding region and the AID in CaN (Figure 2.2). Our data point to a large scale folding of the CaN RD, with ~50 residues adopting α-helical structure, only 24 residues of which are in the CaM binding region. We believe this is the first observation of CaM inducing folding of this scale outside of its binding site on a target protein.

2.5 References

[1] Wang, J.H. & Desai, R. (1976) A brain protein and its effect on the Ca2+-and protein modulator-activated cyclic nucleotide phosphodiesterase, *Biochem. Biophys. Res. Commun. 72*, 926–932

[2] Watterson, D.M. & Vanaman, T.C. (1976) Affinity chromatography purification of a cyclic nucleotide phosphodiesterase using immobilized modulator protein, a troponin C-like protein from brain, *Biochem. Biophys. Res. Commun. 73*, 40–46

[3] Klee, C.B. & Krinks, M.H. (1978) Purification of cyclic 3“,5-”nucleotide phosphodiesterase inhibitory protein by affinity chromatography on activator protein coupled to Sepharose, *Biochemistry 17*, 120–126

[4] Rusnak, F. & Mertz, P. (2000) Calcineurin: form and function, *Physiol. Rev. 80*, 1483–1521

[5] Chakraborti, S., Das, S., Kar, P., Ghosh, B., Samanta, K., Kolley, S., Ghosh, S., Roy, S. & Chakraborti, T. (2007) Calcium signaling phenomena in heart diseases: a perspective, *Mol. Cell Biochem. 298*, 1–40

[6] Vega, R.B., Bassel-Duby, R. & Olson, E.N. (2003) Control of cardiac growth and function by calcineurin signaling, *J. Biol. Chem. 278*, 36981–36984

[7] Ermak, G., Morgan, T.E. & Davies, K.J. (2001) Chronic overexpression of the calcineurin inhibitory gene DSCR1 (Adapt78) is associated with Alzheimer's disease, *J. Biol. Chem. 276*, 38787–38794

[8] Hoeffer, C.A., Dey, A., Sachan, N., Wong, H., Patterson, R.J., Shelton, J.M., Richardson, J.A., Klann, E. & Rothermel, B.A. (2007) The Down syndrome critical region protein RCAN1 regulates long-term potentiation and memory via inhibition of phosphatase signaling, *J. Neurosci. 27*, 13161–13172

[9] Wang, P. & Heitman, J. (2005) The cyclophilins, *Genome Biol. 6*, 226

[10] Klee, C.B., Ren, H. & Wang, X. (1998) Regulation of the calmodulin-stimulated protein phosphatase, calcineurin, *J. Biol. Chem. 273*, 13367–13370

[11] Manalan, A.S. & Klee, C.B. (1983) Activation of calcineurin by limited proteolysis, *Proc. Natl. Acad. Sci. U.S.A 80*, 4291–4295

[12] Stemmer, P.M. & Klee, C.B. (1994) Dual calcium ion regulation of calcineurin by calmodulin and calcineurin B, *Biochemistry 33*, 6859–6866

[13] Yang, S.A. & Klee, C.B. (2000). Low affinity Ca2+-binding sites of calcineurin B mediate conformational changes in calcineurin A, *Biochemistry 39*, 16147–16154

[14] Swulius, M.T. & Waxham, M.N. (2008). Ca(2+)/calmodulin-dependent protein kinases, *Cell Mol. Life Sci. 65*, 2637–2657

[15] Kissinger, C.R., Parge, H.E., Knighton, D.R., Lewis, C.T., Pelletier, L.A., Tempczyk, A., Kalish, V.J., Tucker, K.D., Showalter, R.E. & Moomaw, E.W. (1995) Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex, *Nature 378*, 641–644

[16] Romero, Obradovic & Dunker, K. (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family, *Genome Inf. Workshop on Genome Inf. 8*, 110–124

[17] Shen, X., Li, H., Ou, Y., Tao, W., Dong, A., Kong, J., Ji, C. & Yu, S. (2008) The secondary structure of calcineurin regulatory region and conformational change induced by calcium/calmodulin binding, *J. Biol. Chem. 283*, 11407–11413

[18] Perrino, B.A. (1999) Regulation of calcineurin phosphatase activity by its autoinhibitory domain, *Arch. Biochem. Biophys. 372*, 159–165

[19] Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C. & Obradovic, Z. (2001) Intrinsically disordered protein, *J. Mol. Graph. Model. 19*, 26–59

[20] Zhao, C., Hellman, L.M., Zhan, X., Bowman, W.S., Whiteheart, S.W. & Fried, M.G. (2010) Hexahistidine-tag-specific optical probes for analyses of proteins and their interactions, *Anal. Biochem. 399*, 237–245

[21] Provencher, S.W. & Glöckner, J. (1981) Estimation of globular protein secondary structure from circular dichroism, *Biochemistry 20*, 33–37

[22] Sreerama, N. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, *Anal. Biochem. 287*, 252–260

[23] Keppel, T.R., Jacques, M.E., Young, R.W., Ratzlaff, K.L. & Weis, D.D. (2011) An efficient and inexpensive refrigerated LC system for H/D exchange mass spectrometry, *J. Am. Soc. Mass Spectrom. 22*, 1472–1476

[24] Hubbard, M.J. & Klee, C.B. (1989) Functional domain structure of calcineurin A: mapping by limited proteolysis, *Biochemistry 28*, 1868–1874

[25] Perrino, B.A., Ng, L.Y. & Soderling, T.R. (1995) Calcium regulation of calcineurin phosphatase activity by its B subunit and calmodulin. Role of the autoinhibitory domain, *J. Biol. Chem. 270*, 340–346

[26] Hashimoto, Y., Perrino, B.A. & Soderling, T.R. (1990) Identification of an autoinhibitory domain in calcineurin, *J. Biol. Chem. 265*, 1924–1927

[27] Lakowicz, J.R. (Springer Verlag: 2006), *Principles of fluorescence spectroscopy*, 954

[28] Quintana, A.R., Wang, D., Forbes, J.E. & Neal Waxham, M. (2005) Kinetics of calmodulin binding to calcineurin, *Biochem. Biophys. Res. Commun. 334*, 674–680

[29] O'Donnell, S., Yu, L., Fowler, C.A. & Shea, M.A. (2011) Recognition of β-calcineurin by the domains of calmodulin: Thermodynamic and structural evidence for distinct roles, *Proteins 79*, 765–786

[30] Henkels, C.H. & Oas, T.G. (2006) Ligation-state hydrogen exchange: coupled binding and folding equilibria in ribonuclease P protein, *J. Am. Chem. Soc. 128*, 7772–7781

[31] Keppel, T.R., Howard, B.A. & Weis, D.D. (2011) Mapping unstructured regions and synergistic folding in intrinsically disordered proteins with amide H/D exchange mass spectrometry, *Biochemistry*

[32] Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N. & Dunker, A.K. (2007) Intrinsic disorder and functional proteomics, *Biophys. J. 92*, 1439–1456

[33] Dyson, H.J. & Wright, P.E. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54–60

[34] Radivojac, P., Vucetic, S., O'connor, T.R., Uversky, V.N., Obradovic, Z. & Dunker, A.K. (2006) Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition, *Proteins 63*, 398–410

[35] Charbonneau, H., Hice, R., Hart, R.C. & Cormier, M.J. (1983) Purification of calmodulin by Ca2+-dependent affinity chromatography, *Methods Enzymol. 102*, 17–39

[36] Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. & Klenk, D.C. (1985) Measurement of protein using bicinchoninic acid, *Anal. Biochem. 150*, 76–85

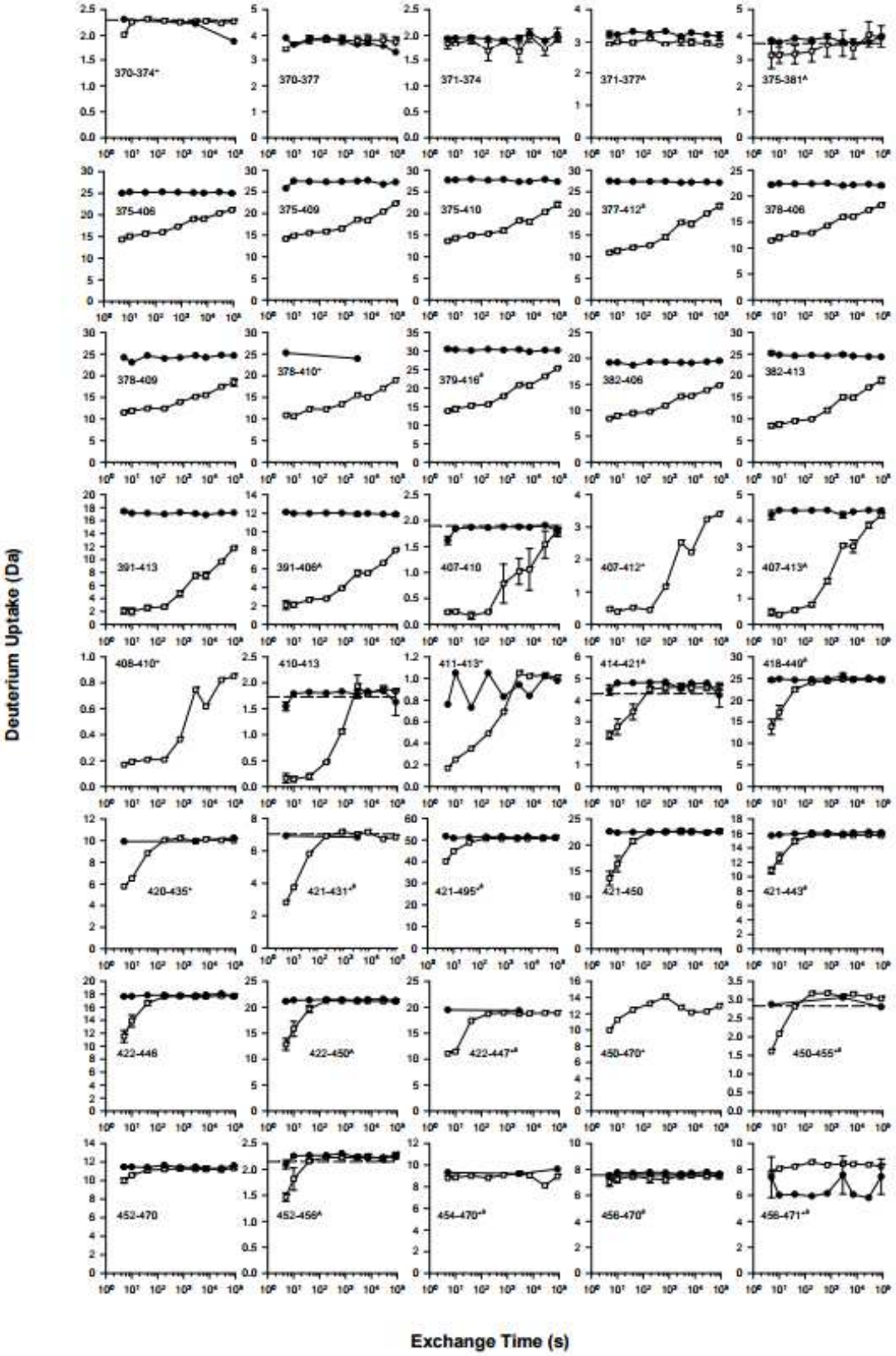
[37] Wang, L., Pan, H. & Smith, D.L. (2002) Hydrogen exchange-mass spectrometry: optimization of digestion conditions, *Mol. Cell Proteom. 1*, 132–138

[38] Busby, S.A., Chalmers, M.J. & Griffin, P.R. (2007) Improving digestion efficiency under H/D exchange conditions with activated pepsinogen coupled columns, *Int. J. Mass Spectrom. 259*, 130–139

[39] Zhang, Z. & Smith, D.L. (1993) Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation, *Protein Sci. 2*, 522-531

Appendices

2.A CaN RD-AID-CT peptide uptake plots



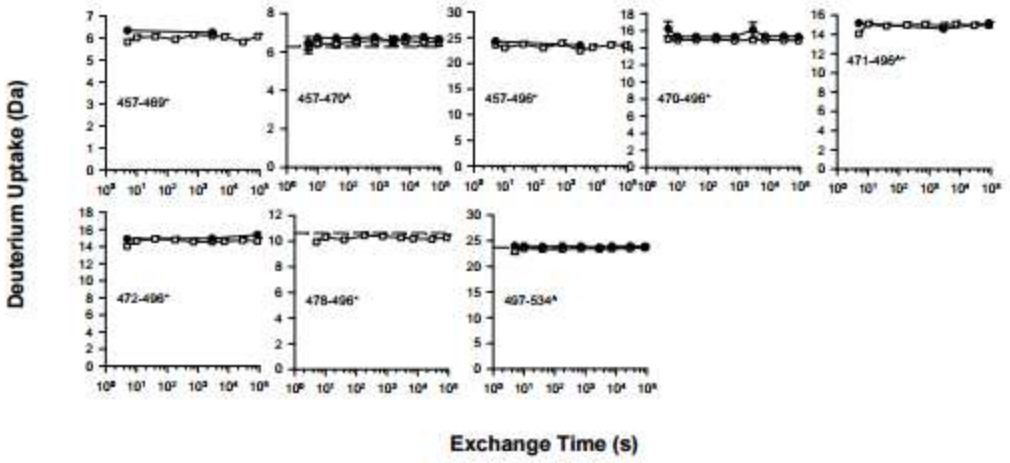


Figure 2.A Deuterium uptake plots for free (filled circles) and CaM-bound (open squares) RD-AID-CT of αCaN. Dashed line shows average deuterium uptake for free RD-AID-CT for cases where totally deuterated peptides were not detected.

2.B Ambiguous (isobaric) peptide assignments

Table 2.B Ambiguous (isobaric) peptide assignments

Residue Numbers	Sequence
377-412	LGSEEDGFDGATAAARKEVIRNKIRAIGKMARVFSV
378-413	GSEEDGFDGATAAARKEVIRNKIRAIGKMARVFSVL
375-413	DELGSEEDGFDGATAAARKEVIRNKIRAIGKMARVFSVL
379-416	SEEDGFDGATAAARKEVIRNKIRAIGKMARVFSVLREE
380-417	EEDGFDGATAAARKEVIRNKIRAIGKMARVFSVLREES
381-418	EDGFDGATAAARKEVIRNKIRAIGKMARVFSVLREESE
418-449	ESVLTLKGLTPTGMLPSGVLSGGKQTLQSATV
419-450	SVLTLKGLTPTGMLPSGVLSGGKQTLQSATVE
421-431	LTLKGLTPTGM
422-432	TLKGLTPTGML
421-495	LTLKGLTPTGMLPSGVLSGGKQTLQSATVEAIEADEAIKGFSPQHKITSF EEAKGLDRINERMPPRRDAMPSDAN
422-496	TLKGLTPTGMLPSGVLSGGKQTLQSATVEAIEADEAIKGFSPQHKITSFE EAKGLDRINERMPPRRDAMPSDANL
421-443	LTLKGLTPTGMLPSGVLSGGKQT
422-444	TLKGLTPTGMLPSGVLSGGKQTL
422-447	TLKGLTPTGMLPSGVLSGGKQTLQSA
423-448	LKGLTPTGMLPSGVLSGGKQTLQSAT

Residue Numbers	Sequence
450-455	EAIEAD
451-456	AIEADE
452-457	IEADEA
453-458	EADEAI
454-470	ADEAIKGFSPQHKITSF
460-476	GFSPQHKITSFEEAKGL
456-470	EAIKGFSPQHKITSF
457-471	AIKGFSPQHKITSFE
456-471	EAIKGFSPQHKITSFE
457-472	AIKGFSPQHKITSFEE
458-473	IKGFSPQHKITSFEEA
461-476	FSPQHKITSFEEAKGL

2.C SDS-PAGE for limited tryptic digestion

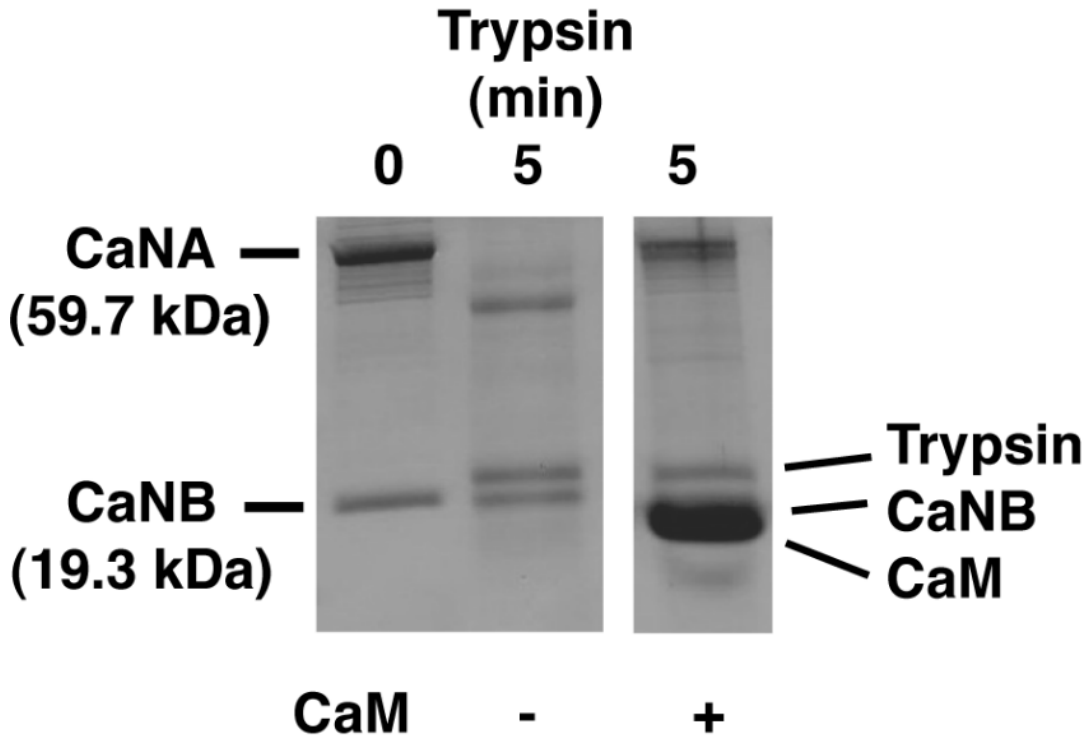
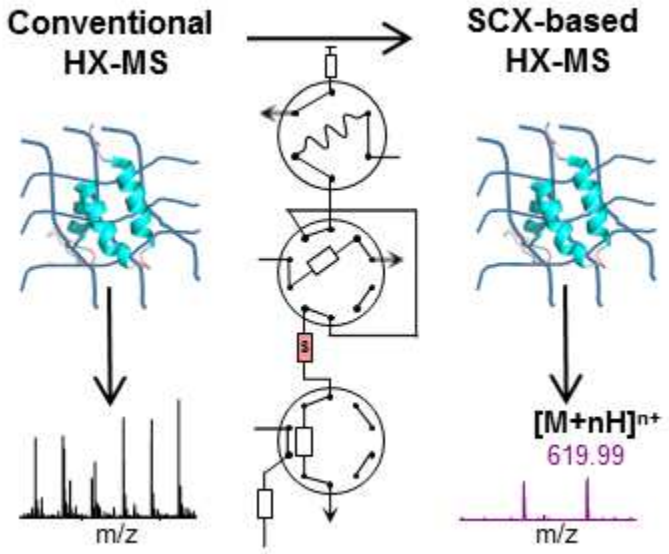


Figure 2.C SDS-PAGE showing results of limited trypic digest of αCaN in the absence and presence of CaM. The αCaN is protected from digestion by addition of CaM

Chapter 3: Automated strong cation exchange cleanup to remove macromolecular crowding agents for protein hydrogen exchange mass spectrometry[†]



[†] Reprinted with permission from Rusinga, F. I., and Weis, D. D. (2017) Automated strong cation exchange cleanup to remove macromolecular crowding agents for protein hydrogen exchange mass spectrometry, *Anal. Chem. 89*, 1275-1282. Copyright (2017) American Chemical Society.

3.1 Introduction

Intrinsically disordered proteins (IDPs) lack stable secondary or tertiary structure and are highly flexible in solution. The degree of disorder varies between well-folded proteins with extended, disordered regions to entirely unfolded proteins.[1] Despite the prevailing protein structure-function paradigm, disorder is an important feature for protein functions.[2] Disordered proteins are involved in many cell signaling and regulatory processes where their flexibility allows some of them to bind many distinct binding partners and adopt multiple conformations.[3, 4] Furthermore, sequences that are predicted to be disordered are very common in proteins that are implicated in diseases such as Alzheimer's and various cancers, making IDPs a medically relevant class of proteins.[5-7] The biophysical properties of IDPs have been characterized mostly under dilute, *in vitro* conditions, yet most IDPs function in cells where the proteins are crowded by various macromolecules.[8] While structured proteins are generally stabilized by macromolecular crowding, less is known about the effects of macromolecular crowding on disordered proteins.

Macromolecular crowding occurs in cells because of the high concentrations of macromolecules that are dissolved in most biological media. For example, the cytoplasm of *E. coli* contains about 300 g L^{-1} of proteins and nucleotides.[9] Eukaryotic cells also contain various networks of extended macromolecular structures and numerous membranes.[9-11] Macromolecular crowding can affect protein folding through volume exclusion effects (i.e. steric repulsions) and non-specific interactions between proteins and other macromolecules sharing a limited space.[12-15] Volume exclusion effects stabilize globular proteins because compact conformations are favored in highly crowded surroundings.[13, 15, 16] On the other hand, the structural plasticity of IDPs

suggests that crowding might influence their structure and dynamics to a greater extent and also less predictably than more stable, structured proteins.

Protein structure and function have been studied under crowded conditions using sensitive spectroscopic techniques like Förster resonance energy transfer (FRET), circular dichroism (CD), and NMR.[17-20] In most crowding experiments, inert, water-soluble, synthetic polymers such as polysaccharides, polyethylene glycol, and polyvinylpyrolidone are used to mimic cellular crowders. Synthetic polymers possess different spectroscopic properties than proteins. Furthermore, these crowders can be acquired with sufficient purity so that impurities do not present problematic interferences at high, physiologically relevant, concentrations added to simulate cellular crowding. FRET and CD spectroscopy have been used to study IDPs under crowded conditions,[18, 20] but these low resolution techniques only probe global conformational effects, not localized (secondary structure) effects. NMR spectroscopy provides the highest resolution for protein structural information from highly crowded samples but protein size and sample preparation requirements limit the applicability of this technique. Furthermore, NMR resonances of IDPs may become unresolvable as disordered regions of the proteins begin to adopt different conformations under crowded conditions.[21]

Hydrogen exchange mass spectrometry (HX-MS) has become a widely-used technique for probing protein structure and dynamics.[22] HX-MS offers higher spatial resolution than many spectroscopic techniques. Moreover, HX-MS is useful for analyzing IDP structure, as we have previously demonstrated by mapping α-helix formation in the disordered regulatory domain of ACTR and structuring of calcineurin upon binding to their respective protein ligands.[23, 24] While HX-MS is valuable for studying IDPs, the presence of a thousand-fold excess of crowding agents complicates the sample matrix and greatly interferes with HX-MS analysis. HX-MS has already

been applied to analyze proteins in other complex matrices, including membrane-bound proteins on phospholipid bilayer nanodiscs,[25] on Langmuir membrane monolayers,[26, 27] and even in mitochondria.[28] A key feature of these experiments is that mass spectrometry measurement was coupled with LC workflows that are capable of removing the interferents after the quench step or with specialized sample handling devices like Langmuir troughs.

Guided by this idea, we have developed a fully-automated, strong cation exchange (SCX) and reversed-phase based extraction method to remove Ficoll, an inert synthetic sucrose polymer cross-linked by epichlorohydrin that is commonly used in studies of macromolecular crowding. We applied the extraction method to analyze the effects of crowding on a disordered domain of the nuclear coactivator domain of CREB binding protein (CBP), a molten globular IDP.[23, 29-31] HX-MS revealed subtle and opposing changes in the conformational stability of CBP, in the presence of Ficoll. We believe this to be the first report of HX-MS analysis of IDP secondary structure in the presence of physiologically relevant amounts of macromolecular crowders.

3.2 Experimental

The preparation of protein and peptide samples is described in the Appendix 3.A.

3.2.1 HX labeling for standards, myoglobin, and CBP

All mixing and injection was performed by a LEAP H/DX PAL robotic system with temperature-controlled compartments (LEAP Technologies, Carrboro, NC) configured with one six-port injection valve and two ten-port switching valves, configured for the SCX workflow, as shown in Figure 3.1.

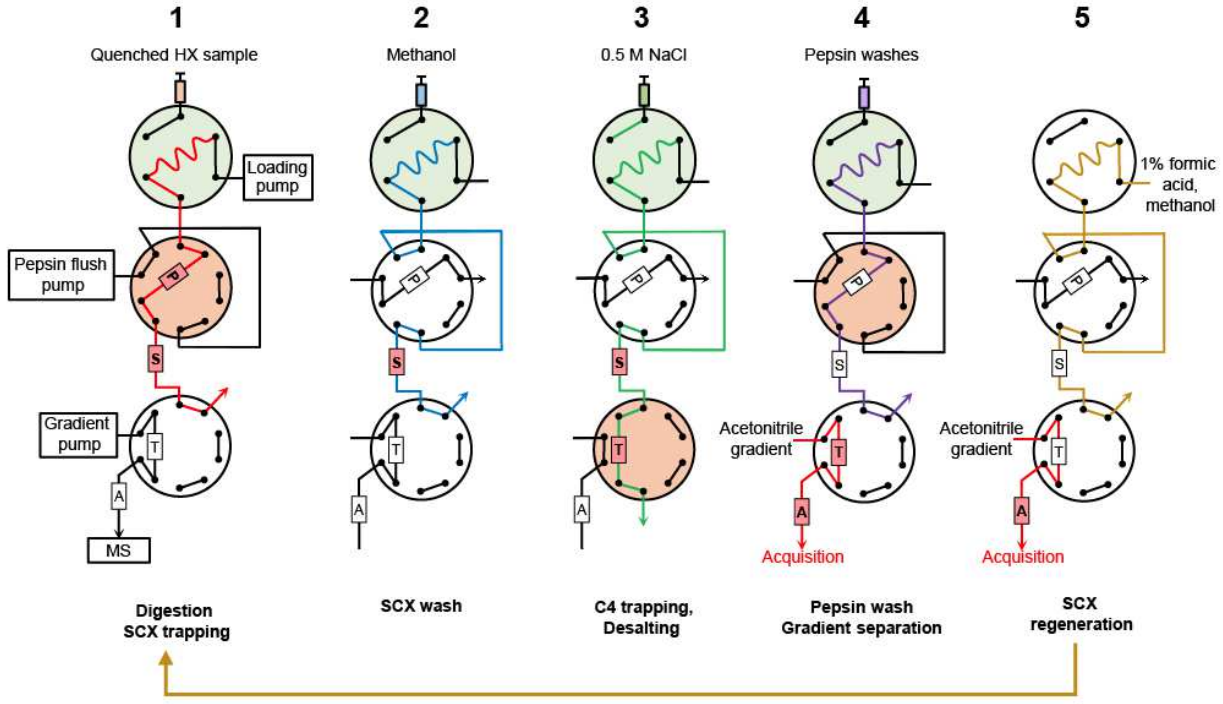


Figure 3.1 SCX cleanup to remove Ficoll for HX-MS analysis of crowded protein samples. The LC components, immobilized pepsin column (P), strong cation exchange column (S), C4 desalting trap (T), and analytical column (A) are assembled in a LEAP H/DX PAL chilled compartment. Injections and valve switches are programmed using LEAP H/DX PAL software.

Deuterated buffers were prepared using pre-deuterated Ficoll (as described in the Appendix *3.A.3*); and are reported as pH adjusted to pD by incorporating the isotope effect.[32] To label, 6 µL of protein or peptide samples (held in a 1 °C compartment) was diluted in 114 µL of citrate buffer in D_2O (pD 4 for control peptides and pD 6.5 for CBP) and then incubated for times ranging from 13 s to 12 h, in a 25 °C labeling compartment. Undeuterated controls were labeled similarly in citrate buffer in H_2O. After incubation, the labeled samples (and undeuterated controls) were quenched by diluting them 1:1 with 200 mM phosphate in H_2O (pH 2.5) in the 1 °C quench compartment.

To measure peptide recoveries from SCX workflow, 10 µM of intact myoglobin and 38 µM CBP were diluted with H_2O citrate buffer, quenched, and loaded into the LC for both

conventional and SCX workflows. As a control for HX kinetics and available deuterium in crowded buffers, peptic peptides were labeled in pD 4 buffer for times between 13 s and 2 h, and loaded into the LC for SCX workflow. Crowded and uncrowded peptide samples were labeled in crowded and uncrowded citrate buffers respectively. The unstructured, predigested peptides were judged to be completely exchanged after 2 hours of HX and thus represent the totally deuterated control for fractional HX calculations. For CBP HX, crowded and uncrowded protein samples (38 μM) were labeled in crowded and uncrowded citrate buffer (pD 6.5), respectively, for times between 13 s and 12 h. The 12 h time represented the totally deuterated control. After labeling and quenching, samples were loaded into LC for SCX workflow.

3.2.2 Chromatography: Conventional and SCX workflow

All LC workflows were performed in a LEAP H/DX PAL LC compartment held at 0 ºC. Valve switches and wash solution injections were programmed in HDXDirector software. For the conventional workflow, the quenched samples were carried through a 50 mm × 2.1 mm immobilized pepsin column by a loading pump (Agilent Technologies 1260 series quaternary pump) at a flow rate of 300 μL min^{-1} of 0.1 % formic acid. The peptic peptides were desalted and concentrated on a self-packed 10 mm × 1 mm C_4 trap (Jupiter, Phenomenex, Torrance, CA). After the desalting step, the trap was placed into the gradient pump flow-path (Agilent Technologies 1260 series binary pump) by switching valve 3. Peptides were eluted from the trap and subsequently separated on a C_{18} column (50 mm × 1mm, Zorbax 300SB-C18, 3.5 μm particles, Agilent Technologies) with 0.1 % formic acid as solvent A and 90 % acetonitrile/10 % water/0.1 % formic acid as solvent B. Peptides were separated over a gradient of 10 to 35 % B over 5 minutes at a flow rate of 50 μL $min^{-1`}$.

For the SCX workflow (see Figure 3.1), the loaded CBP or peptide standard samples were carried through the immobilized pepsin column as in the conventional workflow. The resulting peptic peptides were then trapped on a 10 mm × 2.1 mm strong cation exchange column (BioBasic SCX, Thermo Scientific) while untrapped Ficoll from crowded samples flowed to waste. After 2 minutes of loading and washing, valve 2 was switched so that the pepsin column was washed by the pepsin back-flush pump (Agilent Technologies 1260 series isocratic pump). An injection of 250 µL of chilled methanol with 0.1 % formic acid added was delivered to the SCX column by the loading pump to remove retained Ficoll impurities. After an additional minute, valve 3 was switched and an injection of 250 µL of chilled 0.5 M NaCl in 0.1 % formic acid (pH 2.5) was delivered to the SCX column to elute the peptides. Next, the peptides eluted from the SCX column were desalted and concentrated for 2 minutes on the C_4 trap. Valve 3 was then switched back and the gradient pump was used for separation of peptides on the C_{18} column using the same gradient described for the conventional HX-MS measurement. At the start of the gradient, valve 2 was switched for pepsin column washing to reduce carry-over of peptides.[33] The wash consisted of a mild solvent wash mixture of acetonitrile (5%), 2-propanol (5%), and acetic acid (20%) in water followed by 2 M guanidine hydrochloride in 100 mM phosphate buffer (pH 2.5) both delivered through the pepsin column by the loading pump at 400 µL min^{-1}. Afterwards, the pepsin column was again diverted from the loading pump flow. Finally, 1 % formic acid from the quaternary loading pump flowed through the SCX column for 1 minute, followed by methanol for 5 minutes to regenerate the SCX phase.

3.2.3 Data Acquisition and Analysis

Mass spectrometry acquisition is described in the Supporting Information. HX feature finding and peptide mass determinations were performed using HDExaminer (Sierra Analytics,

Modesto, CA) and validated by manual inspection. The data were then exported to a spreadsheet and further analyzed using an in-house R-script to generate deuterium uptake plots and standard deviations at each time-point for each labeling condition (i.e., crowded versus uncrowded).

The significance of the differences between crowded and uncrowded uptake plots for each peptide was determined by pooling the standard deviations at all time points for each peptide in each condition. A two-sample Student's *t*-test assuming equal variance was then applied to compare the difference in uptake at each time point at the 98 % confidence interval. To minimize the incidence of false positives that can arise in multiple comparisons,[34, 35] the peptides were only classified as significantly different if at least two measured differences (i.e., two HX labeling times) were different using the *t*-test.

The magnitude of the statistically significant differences in deuterium uptake between the two conditions was calculated for each peptide using the following equation:

$$\Delta\overline{\mathrm{HX}} = \frac{\sum_{i=1}^{n} \bar{m}_{\mathrm{c},i} - \bar{m}_{\mathrm{u},i}}{n} \times \frac{1}{\bar{m}_{\infty} - m_0} \tag{1}$$

where $\bar{m}_{\mathrm{c},i}$ and $\bar{m}_{\mathrm{u},i}$ denote the average mass of a peptide under crowded and uncrowded conditions, respectively, within technical replicates of the same HX labeling time; n denotes the number of distinct HX labeling times measured; $\bar{m}_{\infty}$ is the mass obtained from a fully deuterated control at the 12 h time-point for CBP HX or 2 h time-point peptide crowding HX control experiment; and $\bar{m}_0$ is the undeuterated average mass. The right-hand side of (1) is algebraically equivalent to

$$\Delta\overline{\mathrm{HX}} = \frac{\bar{m}_{\mathrm{c}} - \bar{m}_{\mathrm{u}}}{\bar{m}_{\infty} - m_0} \tag{2}$$

where $\bar{m}_c$ and $\bar{m}_u$ are obtained by taking the average mass over all HX labeling times. $\Delta\overline{HX}$ is summed over all the labeling time-points; $\Delta\overline{HX}$ represents the difference between the average deuterium uptake of a peptide across all HX times in crowded and uncrowded buffers, normalized by maximum observable exchange; its theoretical range is between –1 and 1.

Back-exchange in heavily deuterated peptide standards was calculated according to the following equation:

$$\text{back-exchange} = \left(F - \frac{\bar{m}_\infty - m_0}{N} \right) \times 100\% \tag{3}$$

where F is the atom fraction of deuterium after dilution of the protein ($F = 0.95$ here) and N is the number of exchangeable amides in the peptide, obtained by summing non-proline residues starting at the third residue.

3.3 Results and Discussion

3.3.1 SCX-based cleanup for HX-MS of highly crowded samples

Most macromolecular crowding agents would significantly interfere with LC-MS analysis. Despite being highly soluble, Ficoll and its impurities are not adequately removed during the desalting step that is typical of most bottom-up HX-MS analyses. Besides interfering with peptide mass spectra, Ficoll also binds to the desalting trap and needs extensive washing to prevent it from carrying over between LC runs. To perform HX-MS of samples containing 300 g L^{-1} of Ficoll, we developed an automated, rapid strong cation exchange (SCX) and reversed-phase LC extraction method for use under quench conditions (pH 2.5, 0ºC).

For the SCX-based extraction, a pepsin column, SCX column, and C_4 desalting trap, are sequentially connected on a 3-valve LC setup in a compartment held at 0 ^oC shown in Figure 3.1. Under quench conditions (pH 2.5), most peptic peptides are positively charged so that they will be trapped on the SCX column in step 1. Crucially, Ficoll remains neutral under quench conditions so that it will flow through the SCX column to waste. The pepsin column is then bypassed by switching valve 2 while residual Ficoll and contaminants are washed from the SCX column by methanol in step 2. Bypassing the pepsin column prevents methanol from flowing through the pepsin column where it would irreversibly denature the pepsin. Also, injecting the methanol via the sample loop (instead of delivering by the loading pump) makes it is easier to calculate the dwell time and thus avoid inadvertently poisoning the pepsin with methanol over time. Proper timing of the methanol delivery also prevents salt, from the subsequent 0.5 M NaCl injection, from precipitating out of solution in the LC lines.

After the wash of the SCX column with methanol, valve 3 switches to place the C_4 trap in line with the SCX column. The peptides are then eluted from the SCX column onto the C_4 trap using a 0.5 M NaCl injection in step 3. Next, the peptides are concentrated and desalted on the C_4 trap using the loading pump flow. After the desalting step, valve 3 switches back for a water-acetonitrile gradient to elute the peptides from the C_4 trap for chromatographic separation by an analytical column, en route to MS analysis in step 4. Simultaneously, valve 2 switches for pepsin column regeneration by pepsin wash solution injections. Valve 2 then switches again for the SCX column to be regenerated with 1% formic acid and cleaned by methanol from the loading pump in step 5.

3.3.2 Validation with online digestion of myoglobin in Ficoll

To assess the effectiveness of the SCX cleanup, we compared the recovery of myoglobin peptides from samples containing Ficoll using both conventional and SCX workflows. The mass spectrum in Figure 3.2A shows peptides from 10 µM myoglobin stock diluted twenty-fold with H_2O buffer, quenched, and subjected to the conventional LC workflow before MS analysis. The spectrum was selected over the retention time window where Ficoll impurities typically elute. Figure 3.2B is 10 µM myoglobin stock containing 300 g L^{-1} diluted twenty-fold with H_2O buffer containing 300 g L^{-1} Ficoll, quenched, and also analyzed using the conventional LC workflow before MS analysis. As shown in the figure, we observed a complete loss of peptide ions owing to significant ion suppression by Ficoll impurities. For emphasis and to show the low intensity interference that persists over the entire elution gradient, the spectrum in Figure 3.2B is magnified 10×, compared to Figure 3.2A.

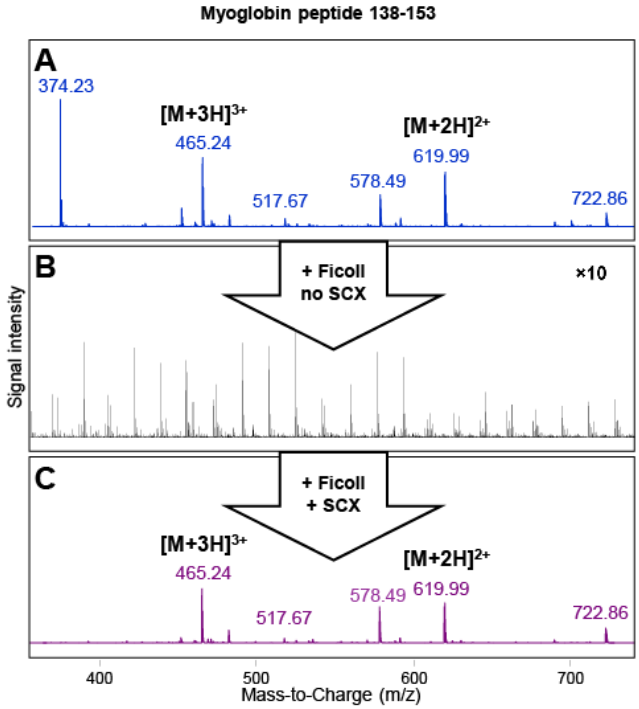


Figure 3.2 Mass spectra summed over the retention time window of a selected myoglobin peptide. (A) The spectrum from a conventional HX-MS experiment. (B) Loss of signal in a Ficoll-containing crowded sample without SCX cleanup. (C) Recovery of peptide spectral features after SCX cleanup. (A) and (C) are on the same scale but (B) is magnified 10× to show low abundance interference.

After applying the SCX workflow to the same crowded myoglobin sample we observed a substantial recovery of peptide spectra and a reduction of interference as shown in Figure 3.2C. Additionally, some relatively high abundance interfering features, such as the ion at m/z = 374, which were not assigned as myoglobin peptides, were eliminated, indicating that SCX cleanup may also remove other co-eluting contaminants from MS analysis. The loss of ion abundance of myoglobin peptide 138-153 in Figure 3.2C was less than 10% compared to that measured in Figure 3.2A.

However, an unavoidable consequence of our cleanup method is the loss of peptide signal intensities due to the increased selectivity introduced by the SCX extraction step. Figure 3.3

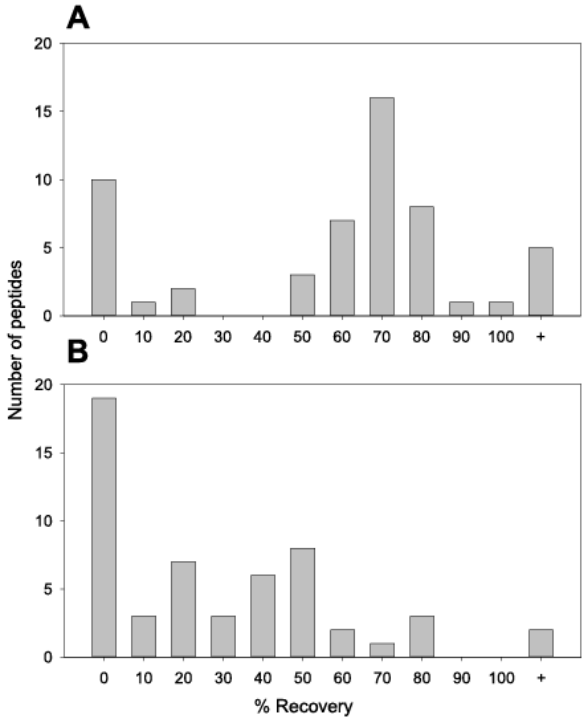


Figure 3.3 Myoglobin peptide recovery from SCX workflow under crowded and non-crowded conditions relative to a conventional HX experiment. (A) Percent recovery based on peptide mass spectral intensity acquired using the SCX workflow on a sample without Ficoll. (B) Percent recovery from a sample containing 300 g L-1 Ficoll.

shows that most myoglobin peptides have reduced intensities after passing through the SCX column. The relative intensities (as a percentage of intensities from the conventional workflow) of most peptide signals after SCX cleanup of uncrowded myoglobin samples, as shown in Figure

However, an unavoidable consequence of our cleanup method is the loss of peptide signal intensities due to the increased selectivity introduced by the SCX extraction step. Figure 3.3 shows that most myoglobin peptides have reduced intensities after passing through the SCX column. The relative intensities (as a percentage of intensities from the conventional workflow) of most peptide signals after SCX cleanup of uncrowded myoglobin samples, as shown in Figure 3.3A, were 60 – 80 %. For crowded samples, as shown in Figure 3.3B, the peptide signal recoveries were further reduced to 20 - 50 % for most peptides. The number of peptide ions that are completely lost (i.e., 0% recovery) was also higher in the crowded samples. Despite the relatively low peptide signal intensities from crowded samples, we still recovered a useful number of overlapping peptides for good HX spatial resolution (see Figure 3.D for the peptide map). Table 3.1 shows that a reasonable number of myoglobin peptides, representing 97% sequence coverage, were recovered from the crowded sample.

Table 3.1 Comparison of number of recovered myoglobin digest peptides and sequence coverage from a conventional HX workflow and from the SCX workflow for samples with and without 300 g L-1 Ficoll.

	Myoglobin		CBP	
	# of peptides	**Sequence coverage (%)**	**# of peptides**	**Sequence coverage (%)**
Conventional HX	54	100	38	100
SCX without Ficoll	44	98	33	100
SCX with Ficoll	35	97	23	100

Although we observed some substantially reduced peptide signal intensities after SCX cleanup from crowded samples, only the lowest abundance peptides would be potentially unusable for hydrogen exchange measurements.

3.3.3 Testing for consistency between crowded and uncrowded labeling buffers

Adding high concentrations of crowding agent could have two undesirable effects on hydrogen exchange. First, adding 300 g L^{-1} of Ficoll to the labeling buffer could alter the fraction of available deuterium atoms because of the many exchangeable hydroxyl groups in polysucrose. Second, the high concentration of crowder potentially imposes other chemical and physical changes to the labeling buffer. To test for changes in the amount of available deuterons and for alterations to HX kinetics in the presence of Ficoll, we labeled pre-digested myoglobin peptides in crowded and uncrowded D_2O buffers at pD 4, quenched the samples, and analyzed them using the SCX workflow. We analyzed short myoglobin peptides which are less likely to

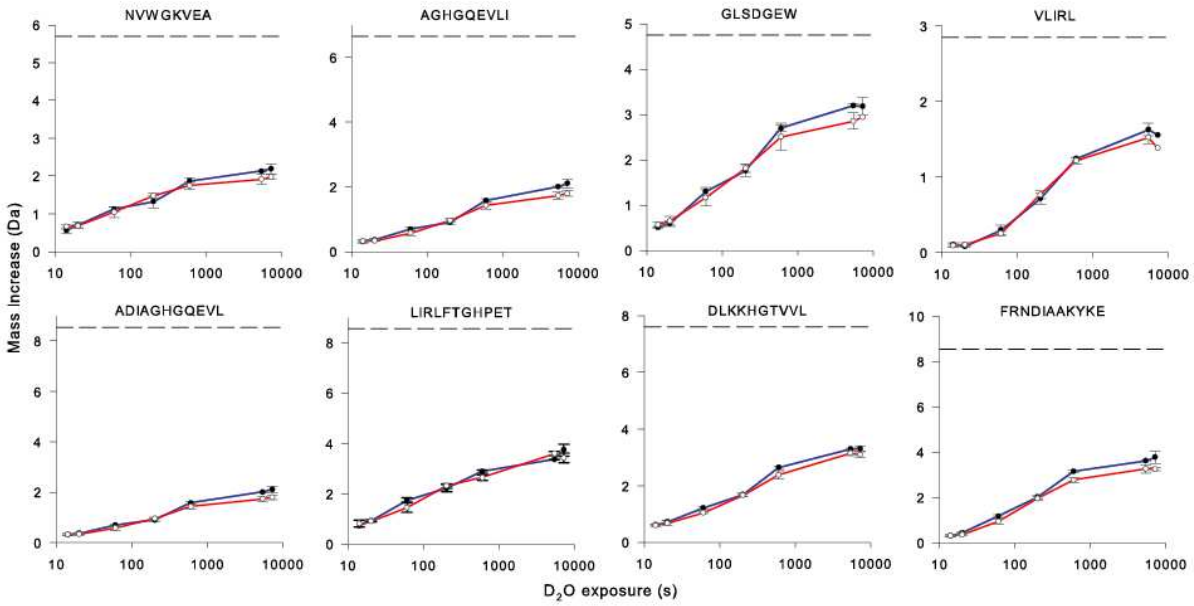


Figure 3.4 Representative deuterium uptake plots for pre-digested myoglobin peptides labeled in crowded (blue, closed circles) and uncrowded (red, open circles) D2O buffers at pD 4. Error bars represent one standard deviation for triplicate data. The overlap shows consistency of HX kinetics in the two buffers. Deuterium uptake also reaches the same maximum level showing that Ficoll does not dilute available deuterium. The dashed lines show the theoretical maximum deuterium uptake, illustrating the substantial loss of deuterium during the SCX workflow.

retain their secondary structure from the intact protein. Because the peptides are also likely to exchange very rapidly at physiological pH due to a lack of structural protection, we slowed down HX by lowering the labeling buffers to pD 4.0. Lowering the pH allowed us to sample complete HX kinetics between 13 s and 2 h. We observed overlap of HX uptake plots across all exchange time-points for pre-digested myoglobin peptides (Figure 3.4).

Since these peptides are expected to be unstructured, their chemical HX would remain unaltered by the presence Ficoll if Ficoll had no effect on the HX process. Our observation therefore demonstrates that Ficoll does not impose any changes on the rate of chemical exchange, a result in agreement with other investigations of the effects of crowding on chemical exchange.[36-38] Thus, according to Linderstrøm-Lang theory,[39] any observed changes in HX rate that we observe in proteins can be attributed to effects of crowding on the dynamics of the protein. We also observed similar deuteration levels for all the peptides after 2 h of labeling, showing that pre-deuterated Ficoll does not significantly dilute the fraction of available deuterium atoms in the crowded labeling buffer.

3.3.4 Deuterium loss from using the SCX workflow

To quantify back-exchange, the loss of deuterium label during analysis, we expressed the deuterium uptake of unstructured peptides after 2 h of labeling as a percentage of the maximum theoretical exchange according to equation (3). We observed between 50 % and 76 % deuterium loss in the pre-digested myoglobin peptides (see Table 3.2). These levels of back-exchange are considerably higher than those from a conventional HX workflow where only 10 % to 25 % loss was observed for similar peptides.[40]

Table 3.2 Back-exchange of highly deuterated myoglobin peptide standards after HX with SCX workflow in the presence or absence of 300 g L–1 Ficoll.

		% back-exchange*	
Residues	**Sequence**	**crowded**	**uncrowded**
(1-7)	GLSDGEW	58	62
(12-19)	NVWGKVEA	59	63
(19-29)	ADIAGHGQEVL	66	68
(22-30)	AGHGQEVLI	71	75
(28-32)	VLIRL	50	56
(29-39)	LIRLFTGHPET	60	59
(60-69)	DLKKHGTVVL	60	63
(138-148)	FRNDIAAKYKE	60	65

*The calculations are based on the deuterium uptake after 2 h of labeling in crowded and uncrowded pD 4 buffer. Calculations account for a maximum of 95 % exchange based on the 1:19 dilution of labeling buffer with sample buffer according to equation (3).

We attribute the additional deuterium loss to interactions between the peptides and the SCX phase, which might catalyze back-exchange in the H_2O LC solvent. Interestingly, Sperry et al. did not observed significant deuterium loss from intact protein after passing through a strong anion exchange column (SAX).[41] However, back-exchange was measured using intact protein which did not bind to the SAX column. In contrast, our measurements included digestion through immobilized pepsin followed by binding of peptic peptides to the SCX column. In spite of the elevated back-exchange, our ability to compare HX trends between the crowded and uncrowded states was unhindered. Since crowded and uncrowded samples undergo the same LC wash steps we expect back-exchange to remain consistent between crowded and uncrowded conditions, as confirmed by the HX measurements described in the preceding section

3.3.5 CBP HX-MS using the SCX workflow

After validating the effectiveness of the SCX-based dual extraction method and testing for consistency in both labeling buffers, we performed HX-MS on the CBP model system. CBP is an intrinsically disordered, molten globular protein with three stable α-helical segments but an ill-defined tertiary structure.[29-31] The naïve expectation is that Ficoll would induce protection in some regions of CBP since crowding is expected to stabilize natively structured proteins.[15-17]

We labeled CBP in crowded and uncrowded D_2O buffers (pD 6.5) for HX-MS and recovered 23 CBP peptides, representing 100 % sequence coverage (see Table 3.S1). There was good peptide overlap across most of the protein sequence (see Figure 3.D for the peptide map). We labeled CBP in both crowded and uncrowded labeling buffers for between 13 s and 12 h, in triplicate. Figure 3.5 shows representative deuterium uptake plots for CBP labeled under

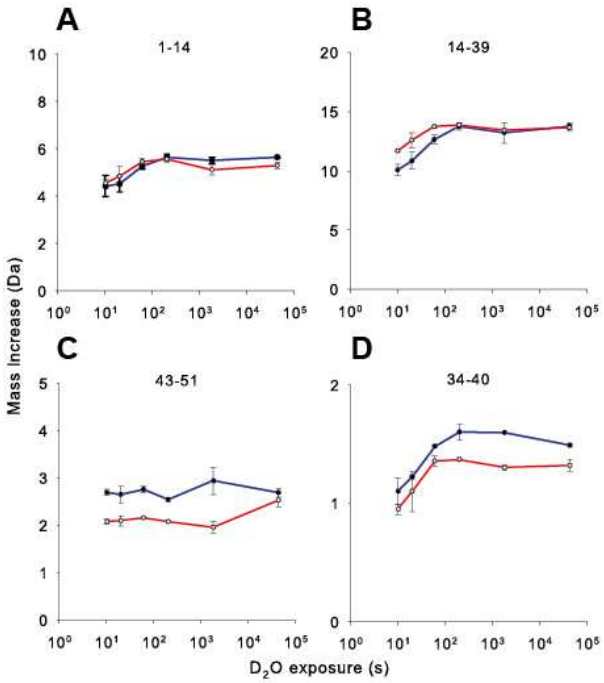


Figure 3.5 Representative deuterium uptake plots for CBP labeled in crowded (blue, closed circles) and uncrowded (red, open circles) D2O buffers. Error bars represent one standard deviation for triplicate measurements; the numbers at the top of each plot denote peptide locations in the CBP sequence. Deuterium uptake plots from (A) a region where there is no significant difference in HX, (B) a region that becomes protected in the presence of Ficoll, (C) a region that becomes substantially deprotected by Ficoll, and (D) a region that becomes slightly deprotected by Ficoll.

crowded (blue, closed circles) and uncrowded (red, open circles) conditions. We observed three general categories of uptake differences between crowded and uncrowded labeling. Figure 3.5A represents the uptake plots for segments that showed no difference in HX between crowded and uncrowded conditions. Figure 3.5B shows a segment where HX was faster in the uncrowded condition. Figure 3.5C and 3.5D show regions where HX was faster under crowded conditions. The extent to which the HX rate is accelerated in crowded conditions is greater in 3.5C than in 3.5D. All uptake differences from the 23 peptides analyzed followed one of the three trends represented by Figures 3.5A-D (see Figure 3.E). Most uptake plots converged at 12 h of HX, further supporting the result, obtained from unstructured peptides, that Ficoll does not alter the amount of deuterium in the labeling buffer.

As we demonstrated in the previous section, Ficoll does not affect the rate of chemical exchange by unstructured peptides, thus any observed differences in uptake are attributable to HX protection or deprotection induced by Ficoll on the protein. Ficoll does not induce any protection in the region covered by peptide 1-14. Peptide 14-39 (see Figure 3.5B) cover a region where HX rates were slowed down by Ficoll crowding, indicating slight protection from HX. Ficoll causes the opposite effect in the region covered by peptide 43-51, where there is an *increase* in HX rates signifying deprotection due to crowding. Finally, Figure 3.5D shows that HX rates also increased in the residues covered by peptide 34-40, although deprotection here is comparatively less than for the region covered by peptide 43-51.

In order to quantify the magnitude of protection or deprotection and thereby describe the effects of crowding with Ficoll on CBP secondary structure, we calculated $\Delta\overline{\mathrm{HX}}$ i.e., the average difference between crowded and uncrowded deuterium uptake across all time-points, as defined by equation (1). Here, $\Delta\overline{\mathrm{HX}} < 0$ indicates Ficoll-induced protection while, $\Delta\overline{\mathrm{HX}} > 0$ indicates

Ficoll-induced deprotection. Figure 3.6 shows $\Delta\overline{HX}$ in each peptide segment, aligned with the primary and secondary structure of CBP. Because there was no significant difference in HX between crowded and uncrowded conditions in all of the peptides from the N-terminal random-coil of CBP, we conclude that this region is unaffected by the presence of Ficoll. Peptides in this region incorporated close to the maximum amount of deuterium after 13s of labeling suggesting that this region is completely unstructured, as has been previously reported for uncrowded CBP.[23, 30]

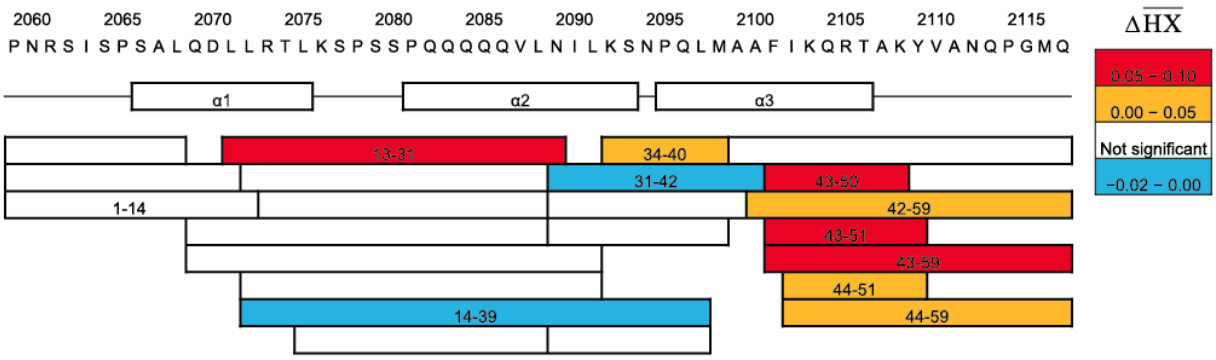


Figure 3.6 CBP protection map showing regions where Ficoll induces protection (blue) and deprotection (red and orange). The degree of protection or deprotection was determined by for peptides that showed statistically significant differences in uptake. was calculated according to equation (1). Only the peptides that showed significant differences and those that are referenced in the text are labeled. Residue 1 in the map corresponds to residue 2059 of the CBP sequence.

Ficoll induced slight protection in residues 14-39 and 31-42 suggesting that this region of CBP is stabilized under highly crowded conditions. We observed similar HX behavior for other peptides in this region, but the differences were not statistically significant (see Figure 3.E). This stabilization by Ficoll is expected because CBP is molten globular protein that is predicted to become more compact based on the theory of volume exclusion.[15] However, the 26 residue peptide (14-39) is too long for us to further localize the residues that are stabilized. In fact,

Ficoll induces slight deprotection in peptides 13-33 and 34-40, which span edges of helical portions but are themselves spanned by the protected peptide, residues 14-39.

Generally, we observed that Ficoll induces deprotection from HX for peptides that are associated with the edges of CBP α-helices. In particular, Figure 3.6 shows that all peptides from α-3 of CBP have positive $\Delta\overline{HX}$ in the presence of Ficoll. Indeed, at least one peptide associated with the other helical segments instead shows slower HX (e.g. 34-40, Figure 3.5D). Ficoll crowding appears to destabilize all CBP α-helices, especially α-3. While this result may seem counter-intuitive, similar observations have been made for α-helical poly(L-glutamic acid) crowded by polyethylene glycol.[18] The tendency for long crowding polymers to form a mesh in solution, instead of acting as individual molecules in good solvent,[42] has been invoked to explain these previously unpredicted observations. The mesh-like network of polymer limits the length of the helical regions by preventing the helix from growing to its maximum possible length. In fact, we have observed similar behavior in another IDP (manuscript in preparation).

3.4 Conclusions

We have successfully removed interference caused by high concentrations of Ficoll using a fully automated SCX-based extraction method for HX-MS analysis of a protein in highly crowded samples. The method can be used to probe the effects of macromolecular crowding on protein structure and dynamics using amide HX. Furthermore, HX-MS with SCX cleanup could be applied to other complex protein samples like those containing nucleic acids, lipids, surfactants, denaturants, and therapeutic protein formulations where the interferents are either negatively-charged or neutral at pH 2-3. Also, while SCX selects peptides because they are positively charged under quench conditions, other extraction phases like strong or weak ion exchange and hydrophobic interaction could be incorporated similarly to select for other HX-MS

analytes (or interferents). To the best of our knowledge, this is the first application of HX-MS to study proteins under highly crowded conditions. In particular, we have been able to identify the effects of crowding on the secondary structure of an IDP, which would be difficult to do by other biophysical methods.

3.5 References

[1] Dunker, A. K., and Obradovic, Z. (2001) The protein trinity—linking function and disorder, *Nat. Biotechnol. 19*, 805-806.

[2] Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, *J. Mol. Biol. 293*, 321-331.

[3] Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) Function and structure of inherently disordered proteins, *Curr. Opin. Struct. Biol. 18*, 756-764.

[4] Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity, *Proc. Natl. Acad. Sci. U.S.A. 93*, 11504-11509.

[5] C Reese, L., and Taglialatela, G. (2011) A role for calcineurin in Alzheimer's disease, *Curr. Neurophys. 9*, 685-692.

[6] Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., and Dunker, A. K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins, *J. Mol. Biol. 323*, 573-584.

[7] Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., and Uversky, V. N. (2007) Functional anthology of intrinsic disorder. III. Ligands, postranslational modifications and diseases associated with intrinsically disordered proteins, *J. Proteom. Res. 6*, 1917-1932.

[8] Theillet, F.-X., Binolfi, A., Frembgen-Kesner, T., Hingorani, K., Sarkar, M., Kyne, C., Li, C., Crowley, P. B., Gierasch, L., and Pielak, G. J. (2014) Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs), *Chem. Rev. 114*, 6661-6714.

[9] Zimmerman, S. B., and Minton, A. P. (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences, *Annu. Rev. Biophys. Biomol. Struct. 22*, 27-65.

[10] Fulton, A. B. (1982) How crowded is the cytoplasm? Minireviews, *Cell 30*, 345-347.

[11] Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli, *J. Mol. Biol. 222*, 599-620.

[12] Gershon, N. D., Porter, K. R., and Trus, B. L. (1985) The cytoplasmic matrix: its volume and surface area and the diffusion of molecules through it, *Proc. Natl. Acad. Sci. U.S.A. 82*, 5030-5034.

[13] Elcock, A. H. (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment, *Curr. Opin. Struct. Biol. 20*, 196-206.

[14] Minton, A. P. (1983) The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences, *Mol. Cell. Biochem. 55*, 119-140.

[15] Zhou, H.-X., Rivas, G., and Minton, A. P. (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences, *Annu. Rev. Biophys. 37*, 375-397.

[16] Sarkar, M., Lu, J., and Pielak, G. J. (2014) Protein crowder charge and protein stability, *Biochemistry 53*, 1601-1606.

[17] Dhar, A., Samiotakis, A., Ebbinghaus, S., Nienhaus, L., Homouz, D., Gruebele, M., and Cheung, M. S. (2010) Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding, *Proc. Natl. Acad. Sci. U.S.A. 107*, 17586-17591.

[18] Koutsioubas, A., Lairez, D., Combet, S., Fadda, G., Longeville, S., and Zalczer, G. (2012) Crowding effect on helix-coil transition: beyond entropic stabilization, *J. Phys. Chem. 136*, 215101-215108.

[19] Miklos, A. C., Li, C., and Pielak, G. J. (2009) Using NMR-detected backbone amide 1 H exchange to assess macromolecular crowding effects on globular-protein stability, *Methods Enzymol. 466*, 1-18.

[20] Soranno, A., Koenig, I., Borgia, M. B., Hofmann, H., Zosel, F., Nettels, D., and Schuler, B. (2014) Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments, *Proc. Natl. Acad. Sci. U.S.A. 111*, 4874-4879.

[21] Wang, Y., Li, C., and Pielak, G. J. (2010) Effects of proteins on protein diffusion, *J. Am. Chem. Soc. 132*, 9392-9397.

[22] Pirrone, G. F., Iacob, R. E., and Engen, J. R. (2015) Applications of hydrogen/deuterium exchange MS from 2012 to 2014, *Anal. Chem. 87*, 99-118.

[23] Keppel, T. R., Howard, B. A., and Weis, D. D. (2011) Mapping unstructured regions and synergistic folding in intrinsically disordered proteins with amide H/D exchange mass spectrometry, *Biochemistry 50*, 8722-8732.

[24] Rumi-Masante, J., Rusinga, F. I., Lester, T. E., Dunlap, T. B., Williams, T. D., Dunker, A. K., Weis, D. D., and Creamer, T. P. (2012) Structural basis for activation of calcineurin by calmodulin, *J. Mol. Biol. 415*, 307-317.

[25] Hebling, C. M., Morgan, C. R., Stafford, D. W., Jorgenson, J. W., Rand, K. D., and Engen, J. R. (2010) Conformational analysis of membrane proteins in phospholipid bilayer nanodiscs by hydrogen exchange mass spectrometry, *Anal. Chem. 82*, 5415-5419.

[26] Pirrone, G. F., Emert-Sedlak, L. A., Wales, T. E., Smithgall, T. E., Kent, M. S., and Engen, J. R. (2015) The membrane-associated conformation of HIV-1 Nef investigated with hydrogen exchange mass spectrometry at a Langmuir monolayer, *Anal. Chem. 87*, 7030-7035.

[27] Pirrone, G. F., Vernon, B. C., Kent, M. S., and Engen, J. R. (2015) Hydrogen exchange mass spectrometry of proteins at Langmuir monolayers, *Anal. Chem. 87*, 7022-7029.

[28] Rey, M., Forest, E., and Pelosi, L. (2012) Exploring the conformational dynamics of the bovine ADP/ATP carrier in mitochondria, *Biochemistry 51*, 9727-9735.

[29] Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators, *Nature 415*, 549-553.

[30] Ebert, M.-O., Bae, S.-H., Dyson, H. J., and Wright, P. E. (2008) NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR†, *Biochemistry 47*, 1299-1308.

[31] Kjaergaard, M., Teilum, K., and Poulsen, F. M. (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP, *Proc. Natl. Acad. Sci. U.S.A. 107*, 12535-12540.

[32] Glasoe, P. K., and Long, F. A. (1960) Use of glass electrodes to measure acidities in deuterium oxide1,2, *J. Phys. Chem. 64*, 188-190.

[33] Majumdar, R., Manikwar, P., Hickey, J. M., Arora, J., Middaugh, C. R., Volkin, D. B., and Weis, D. D. (2012) Minimizing carry-over in an online pepsin digestion system used for the H/D exchange mass spectrometric analysis of an IgG1 monoclonal antibody, *J. Am. Soc. Mass. Spectrom. 23*, 2140-2148.

[34] Toothaker, L. E. (1993) *Multiple comparison procedures*, SAGE Publications, Newbury Park, CA.

[35] Munroe, R. (2016) Significant, xkcd, https://xkcd.com/882/.

[36] Benton, L. A., Smith, A. E., Young, G. B., and Pielak, G. J. (2012) Unexpected effects of macromolecular crowding on protein stability, *Biochemistry 51*, 9773-9775.

[37] Lim, W. K., Rösgen, J., and Englander, S. W. (2009) Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group, *Proc. Natl. Acad. Sci. U.S.A. 106*, 2595-2600.

[38] Wang, A., Robertson, A. D., and Bolen, D. W. (1995) Effects of a naturally occurring compatible osmolyte on the internal dynamics of ribonuclease A, *Biochemistry 34*, 15096-15104.

[39] Hvidt, A., and Nielsen, S. O. (1966) Hydrogen exchange in proteins, *Adv. Protein Chem. 21*, 287-386.

[40] Keppel, T. R., Jacques, M. E., Young, R. W., Ratzlaff, K. L., and Weis, D. D. (2011) An efficient and inexpensive refrigerated LC system for H/D exchange mass spectrometry, *J. Am. Soc. Mass. Spectrom. 22*, 1472-1476.

[41] Sperry, J. B., Wilcox, J. M., and Gross, M. L. (2008) Strong anion exchange for studying protein–DNA interactions by H/D exchange mass spectrometry, *J. Am. Soc. Mass. Spectrom. 19*, 887-890.

[42] de Gennes, P.-G. (1979) *Scaling concepts in polymer physics*, Cornell University Press, Ithaca, NY.

Appendices

3.A Materials and Methods

3.A.1 Protein expression and purification

The nuclear coactivator binding domain of CREB binding protein, residues 2059 – 2117, (UniProt CBP_MOUSE), hereafter referred to as CBP, was expressed in *E. coli* and purified using strong cation exchange as previously described.[1-2] The procedure was modified by buffer exchanging strong cation exchange fractions (from the purification protocol) containing CBP into ultrapure water using a desalting column (HiPrep 26/10, GE Healthcare Life Sciences, Pittsburgh, PA). After purification, the CBP concentration was 17 μM by BCA assay (Thermo Scientific, West Palm Beach, FL) using BSA standards. To concentrate CBP, the stock solution was freeze-dried using a benchtop freeze drier (FreeZone, Labconco, Kansas City, MO). The protein was then reconstituted in 20 mM sodium citrate, 50 mM sodium chloride, pH 6.5 buffer at one-third of the initial volume. The concentrated stock solution contained 38 μM CBP by BCA assay using BSA standards. The stock was divided into aliquots, flash frozen on liquid N_2, and stored at −80 ºC. Crowded CBP samples were prepared by dissolving solid Ficoll PM 70 (Sigma, St. Louis, MO) in thawed aliquots of concentrated CBP stock to reach 300 g L^{-1} of Ficoll.

3.A.2 Myoglobin peptide standards

Peptic peptides of horse heart myoglobin (Sigma) were used to compare deuterium uptake from crowded and uncrowded labeling buffers after the Ficoll removal workflow. Peptides were prepared by passing 10 μM myoglobin in 0.1% formic acid (Thermo Scientific) through a 50 mm × 2.1 mm immobilized pepsin column prepared in-house as previously

described.[3-4] The resulting peptide solution was aliquoted into tubes containing about 3,650 pmol of peptides, dried using a CentriVap system (Labconco), and stored at −20 ºC. For HX-MS, the peptides were reconstituted in 500 µL of citrate buffer (pH 4.0) with and without 300 g L^{-1} Ficoll to make 7.3 µM peptides in crowded and uncrowded standards, respectively.

3.A.3 Pre-deuterated Ficoll for crowded HX labeling buffers

Labeling buffers were prepared with 20 mM sodium citrate, 50 mM sodium chloride in D_2O (pD 6.5, corrected for the isotope effect[33]). Crowded labeling buffer was the same citrate buffer with 300 g L^{-1} of pre-deuterated Ficoll. Ficoll contains numerous labile hydroxyl groups which would dilute the deuterons from the D_2O labeling buffer. To prevent the dilution of deuterons by Ficoll, we pre-deuterated Ficoll by incubating 1.0 g of Ficoll in 10 mL of D_2O for 24 hours at room temperature. After incubation, the Ficoll solution was freeze dried (FreeZone, Labconco) and then stored in a desiccator. ^1H NMR spectra showed that 98 % of the hydroxyl groups in Ficoll became deuterated after incubating in D_2O for 24 hours (see Figure 3.B). There was no further deuteration after subsequent rounds of incubation in D_2O indicating that 24 hrs was sufficient to titrate all exchangeable ^1H atoms (results not shown).

3.A.4 Mass Spectrometry

Peptic peptides were identified using accurate mass measurements (± 10 ppm) and MS/MS using collision-induced dissociation on a quadrupole time-of-flight mass spectrometer (Q-TOF, Agilent model 6530, Santa Clara, CA). Peptide MS and MS/MS spectra were assigned and analyzed using MassHunter Qualitative Analysis software (version B.7.0.0). A set of 23 peptides was found in common between crowded and uncrowded CBP samples (see Table 3.C). Co-eluting peptic peptides sharing a C-terminal cleavage site were further reviewed to prevent

the inclusion of in-source *y* ion fragments of longer peptides: shorter co-eluting peptides sharing a C-terminal cleavage site with longer peptides were rejected as possible in-source fragments unless the N-terminal cleavage site could be matched to another peptide.

3.A.5 References

1. S. J. Demarest, M. Martinez-Yamout, J. Chung, H. Chen, W. Xu, H. J. Dyson, R. M. Evans, P. E. Wright, Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **2002,** *415*, 549-553, DOI: 10.1038/415549a.
2. T. R. Keppel, B. A. Howard, D. D. Weis, Mapping unstructured regions and synergistic folding in intrinsically disordered proteins with amide H/D exchange mass spectrometry. *Biochemistry* **2011,** *50*, 8722-32, DOI: 10.1021/bi200875p.
3. L. Wang, H. Pan, D. L. Smith, Hydrogen exchange-mass spectrometry: optimization of digestion conditions. *Mol. Cell. Proteomics* **2002,** *1*, 132-138.
4. S. A. Busby, M. J. Chalmers, P. R. Griffin, Improving digestion efficiency under H/D exchange conditions with activated pepsinogen coupled columns. *Int. J. Mass Spectrom.* **2007,** *259*, 130-139.

3.B ^{1}H NMR spectra comparing Ficoll and predeuterated Ficoll

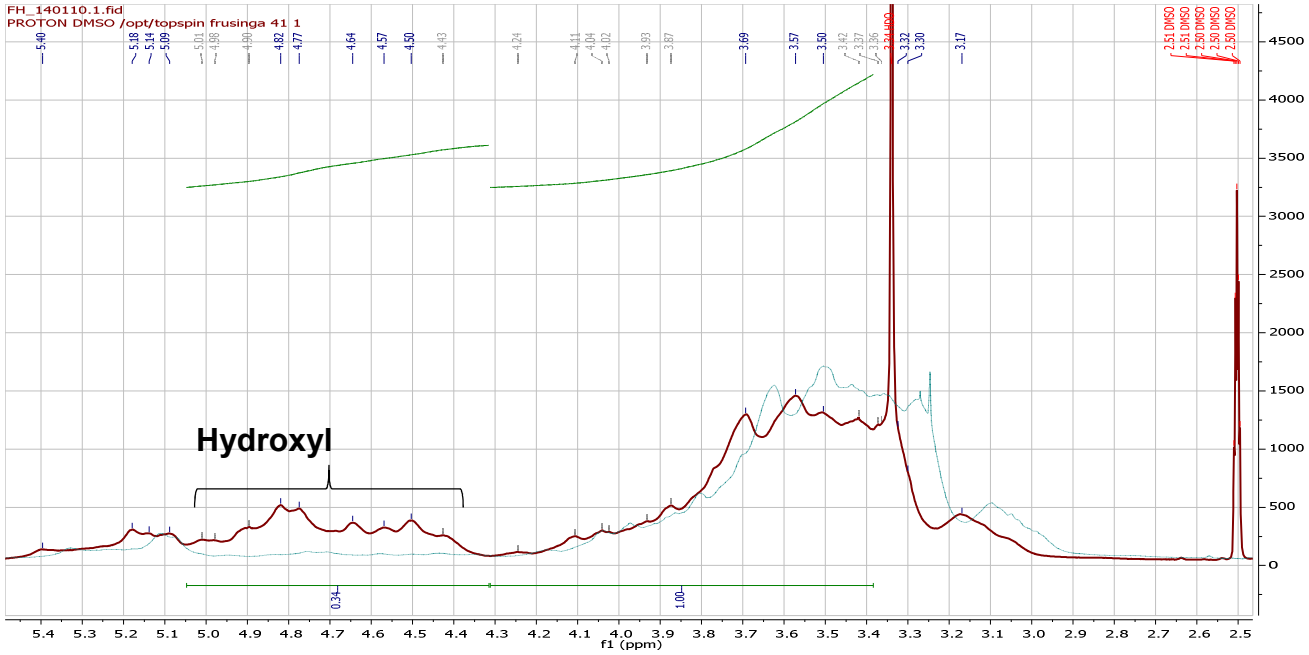


Figure 3.B Overlapped 1H-NMR spectra of undeuterated Ficoll (red) and deuterated Ficoll (blue) showing loss of hydroxyl proton signals after exchanging with deuterons. Maximum exchange occurred after 24 hours of incubation in D2O. No further exchange occurred after incubating for 48 hours (spectrum not shown).

3.C CBP peptide assignments

Table 3.C CBP peptide assignments in common between crowded and uncrowded conditions

Peptide	Sequence	RT	Observed m/z	z	Mass (observed)	Mass (calculated)	Mass error (ppm)
A(1-10)	PNRSISPSAL	7.619	521.29	2	1040.5655	1040.5615	3.92
A(1-13)	PNRSISPSALQDL	8.618	699.3739	2	1396.7335	1396.731	1.79
A(1-14)	PNRSISPSALQDLL	9.506	755.9141	2	1509.8151	1509.8151	-0.03
A(11-30)	QDLLRTLKSPSSPQQQQQVL	8.374	765.4172	3	2293.2317	2293.239	-3.19
A(11-33)	QDLLRTLKSPSSPQQQQQVLNIL	9.25	878.8193	3	2633.4383	2633.4501	-4.46
A(13-31)	LLRTLKSPSSPQQQQQVLN	8.782	722.4035	3	2164.1896	2164.1964	-3.17
A(14-30)	LRTLKSPSSPQQQQQVL	7.504	646.698	3	1937.0718	1937.0694	1.24
A(14-33)	LRTLKSPSSPQQQQQVLNIL	8.871	760.099	3	2277.2758	2277.2805	-2.06
A(14-39)	LRTLKSPSSPQQQQQVLNILKSNPQL	8.842	737.1657	4	2944.6338	2944.6458	-4.09
A(15-30)	RTLKSPSSPQQQQQVL	7.267	609.0021	3	1823.985	1823.9854	-0.2
A(17-30)	LKSPSSPQQQQQVL	7.388	523.2858	3	1566.8342	1566.8366	-1.49
A(31-39)	NILKSNPQL	7.852	513.8022	2	1025.5895	1025.5869	2.51
A(31-40)	NILKSNPQLM	8.17	579.3207	2	1156.627	1156.6274	-0.33
A(31-41)	NILKSNPQLMA	8.115	614.839	2	1227.6635	1227.6645	-0.88
A(31-42)	NILKSNPQLMAA	8.128	650.3574	2	1298.7007	1298.7017	-0.75
A(34-40)	KSNPQLM	7.024	409.2162	2	816.4173	816.4164	1.16
A(41-59)	AAFIKQRTAKYVANQPGMQ	7.328	531.2849	4	2121.112	2121.1153	-1.59
A(42-59)	AFIKQRTAKYVANQPGMQ	7.253	513.5266	4	2050.0776	2050.0782	-0.28
A(43-50)	FIKQRTAK	5.143	331.2064	3	990.5984	990.5975	0.94
A(43-51)	FIKQRTAKY	6.486	289.4235	4	1153.6643	1153.6608	3.08
A(43-59)	FIKQRTAKYVANQPGMQ	7.021	495.7678	4	1979.0419	1979.0411	0.38
A(44-51)	IKQRTAKY	5.737	336.5388	3	1006.5948	1006.5924	2.4
A(44-59)	IKQRTAKYVANQPGMQ	6.687	459.0009	4	1831.9737	1831.9727	0.53

23 CBP peptides were recovered from both crowded and uncrowded samples after SCX workflow, and subsequently used for HX-MS analysis. Residue 1 in the table corresponds to residue 2059 in the CBP sequence.

3.D Myoglobin and CBP peptic peptide maps

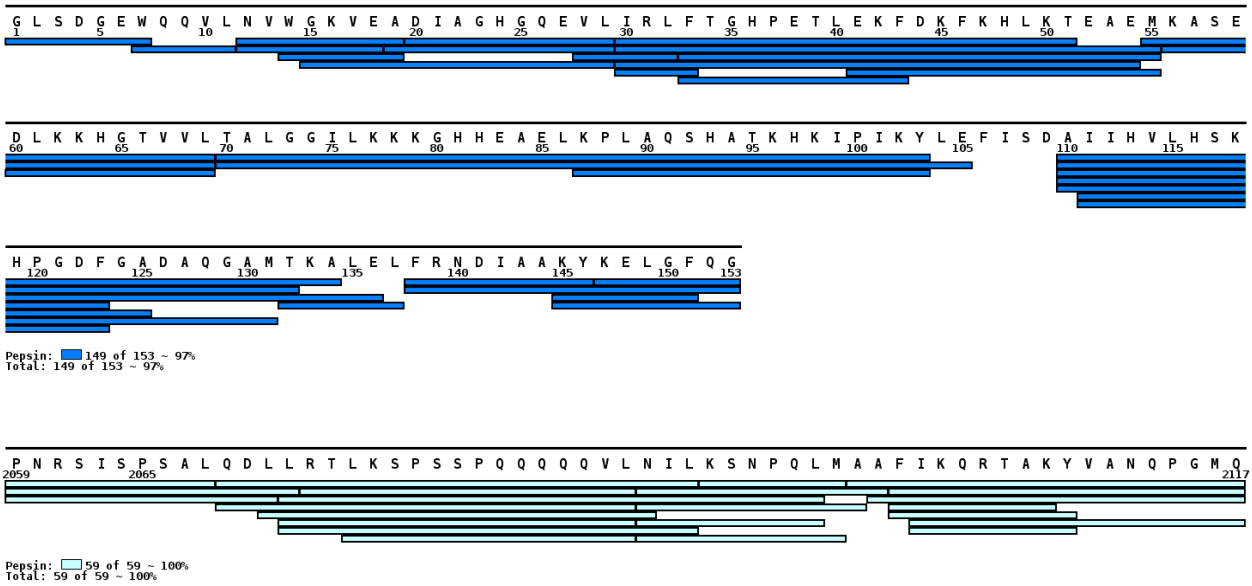
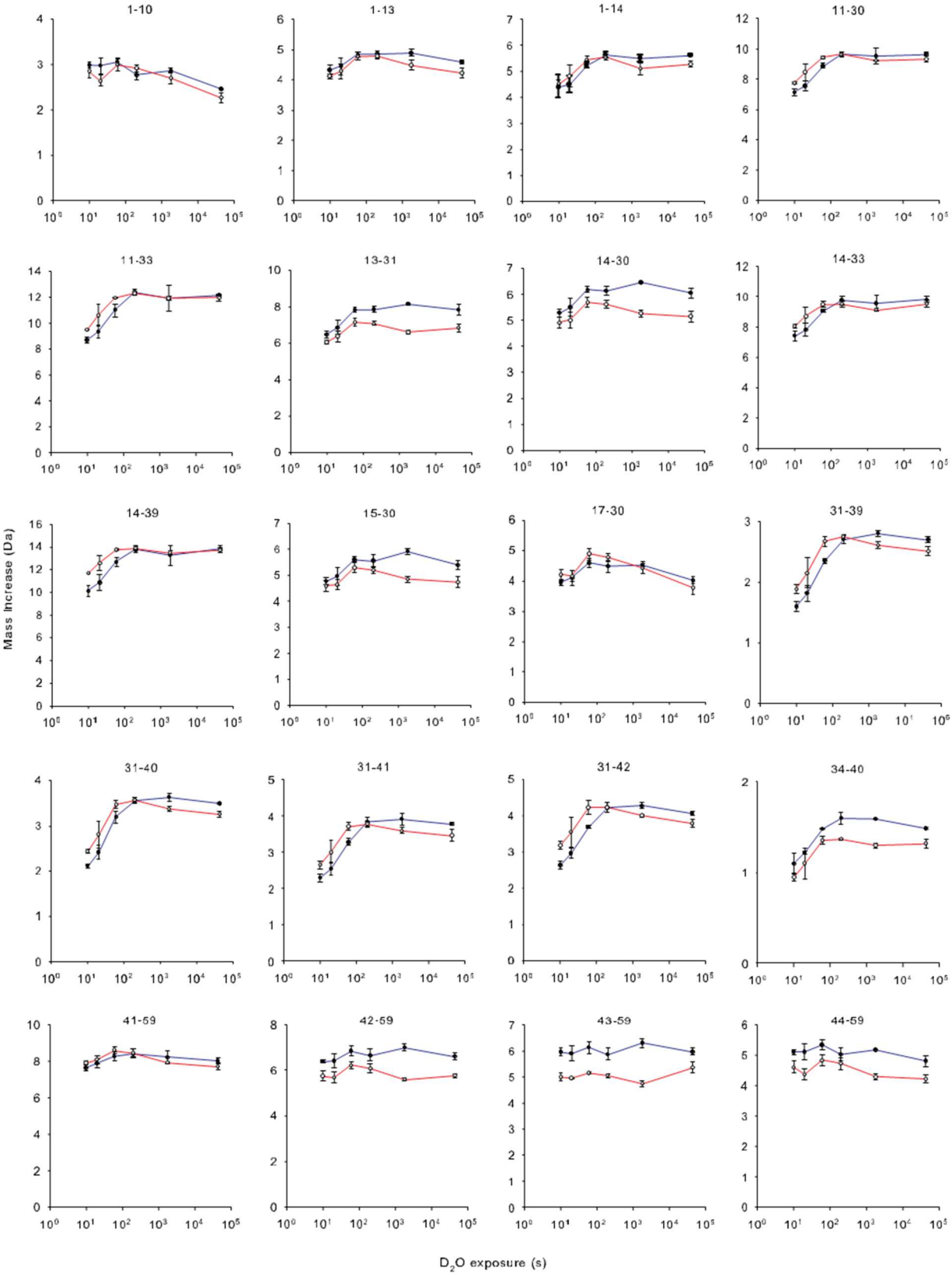


Figure 3.D Myoglobin (blue) and CBP (green) peptic peptide maps from crowded samples after SCX workflow. Myoglobin and CBP samples containing 300 g L^{-1} Ficoll were diluted in H2O buffer with 300 g L^{-1} Ficoll, quenched, and injected into LC for SCX workflow. 23 peptides, covering 100% of the CBP sequence were recovered and used for HX-MS analysis. Maps were generated using MS Tools [Kavan, D. and Man, P. "MSTools - Web based application for visualization and presentation of HXMS data" Int. J. Mass Spectrom. 2011, 302: 53-58. http://dx.doi.org/10.1016/j.ijms.2010.07.030.]

3.E CBP peptide deuterium uptake curves



3.E continued

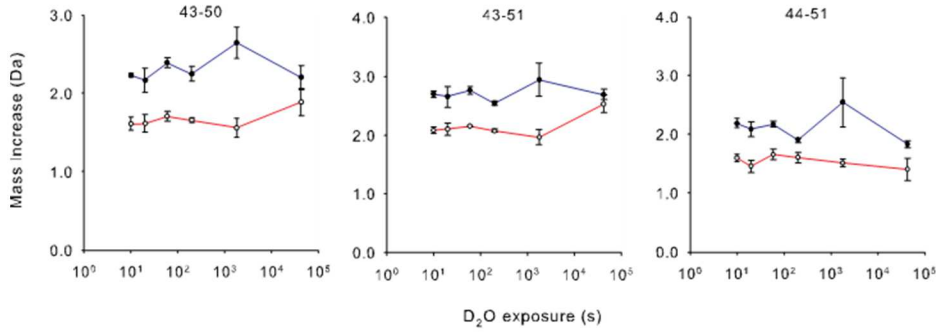
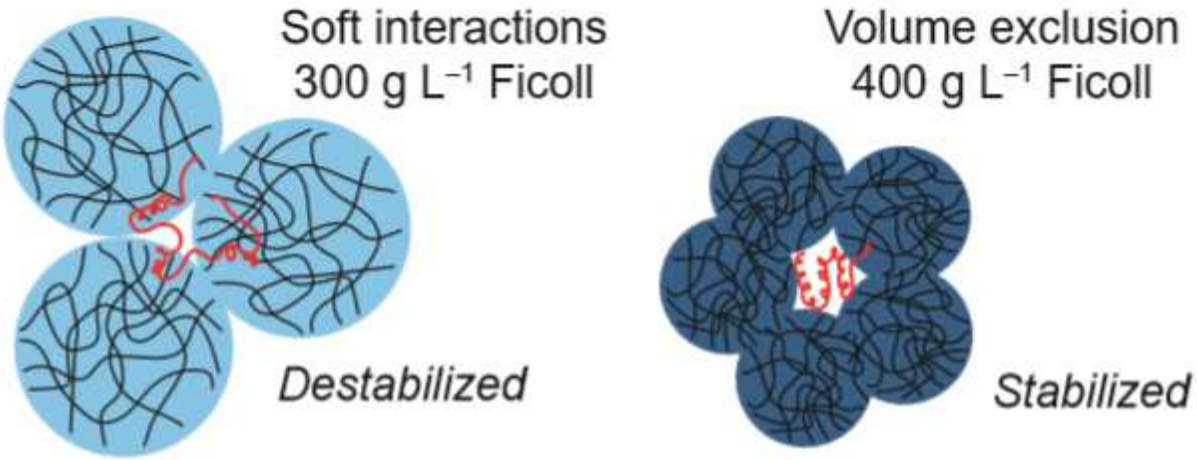


Figure 3.E Deuterium uptake plots for all assigned CBP peptides labeled in crowded (blue, closed circles) and uncrowded (red, open circles) D_2O buffers. Crowded D_2O buffer contained 300 g L^{-1} pre-deuterated Ficoll.

Chapter 4: Soft interactions and volume exclusion by polymeric crowders can stabilize or destabilize transient structure in disordered proteins depending on polymer concentration



4.1 Introduction

Intrinsically disordered proteins (IDPs) are highly flexible in the dilute, *in vitro* solutions in which they are normally studied, but IDP conformational dynamics could be very different in the crowded cellular interiors where they function.[1] Cells contain about 300 g L^{-1} of biological macromolecules[2] that give rise to macromolecular crowding effects. In general, macromolecular crowding stabilizes globular proteins,[3, 4] but crowding effects on IDPs are difficult to predict because IDPs lack a thermodynamically stable folded state. Rather, IDPs populate an ensemble of unstructured conformations. Therefore various interaction between proteins and crowding agents in the crowded environment could affect IDPs in profoundly different ways than structured proteins.

Hard and soft interactions between protein and crowder in the crowded environment can alter protein conformations. Hard interactions, also called volume exclusion, occur because the volume that is occupied by crowding macromolecules is inaccessible to the protein of interest. Hard interactions favor compact protein conformations that occupy less solvent space making unfolded states less entropically favorable[3, 5] Soft interactions are non-specific chemical interactions such as dispersion forces, electrostatic forces, and hydrogen bonding that occur between proteins and crowding agents. Soft interactions depend on the specific chemical properties of the crowder-protein system. These interactions can be stabilizing or destabilizing. Hard and soft interactions are always present, but their relative influences on protein conformations depend on the concentration and solution properties of the crowding agent. At physiologically relevant concentrations, synthetic polymer crowding agents can form mesh-like networks.[6] Thermodynamically, hard interactions drive all proteins towards compact structures,

but IDPs can also become kinetically trapped in a mesh network of polymers that can affect IDP transient structural stability.[7, 8]

In general, protein stability can be understood in terms of the equilibrium proportion of structured states in relation to unstructured states. This proportion can be determined by measuring the rate of backbone amide hydrogen exchange (HX) and applying the Linderstrøm-Lang theory (Scheme 4.1) where k with subscripts denote rate constants for opening (op), closing (cl), and exchange (ch) by amide hydrogens leading to an observed (obs) rate of exchange under the EX2 limit.[9]

$$\mathrm{NH(closed)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{NH(open)} \xrightarrow[D_2O]{k_{ch}} \mathrm{ND}$$

$$k_{obs} = \frac{k_{op}}{k_{cl}} k_{ch} \quad \mathrm{(EX2\ limit)}$$

Scheme 4.1 Linderstrom-Lang scheme

Amide protons that participate in hydrogen bonding or that experience little solvent exposure in hydrophobic cores of proteins only briefly become exchange-competent when the protein unfolds, locally or globally, leading to slow exchange $\left(k_{cl} >> k_{op}\right)$. In contrast, IDP conformational ensembles are highly dynamic so that any protective amide hydrogen bonds exist only briefly such that HX occurs rapidly at most backbone amides $\left(k_{cl} << k_{op}\right)$. Still, HX-MS is sensitive enough to detect transient structure that can form even in random coil IDPs.[10] Macromolecular crowding may alter the folding and unfolding rates of IDPs thereby changing the HX rates of backbone amides compared to IDPs in dilute solutions.

Here, we have used a transiently-helical intrinsically disordered domain of the activator of thyroid and retinoid receptor (ACTR, UniProt NCOA3_HUMAN, residues 1023-1093) as our model IDP, hereafter referred to simply as ACTR, to examine the effects of Ficoll crowding on transient helicity of an IDP. We used HX measured by mass spectrometry (HX-MS) to compare hydrogen exchange kinetics of ACTR in the absence and presence of Ficoll 70, as a crowding agent. Ficoll is an inert, synthetic polysucrose that is highly crosslinked by epichlorohydrin (see Figure 4.1) with an average molecular weight of 70 kDa. Ficoll molecules are spherical, but highly flexible in solution.[11, 12] Ficoll has previously been recommended for mimicking cellular crowding agents.[3] At physiologically relevant concentrations, Ficoll is known to stabilize the folded structures of globular proteins,[13] but the effects of Ficoll on IDP structure have only been investigated in a few studies.[14, 15]

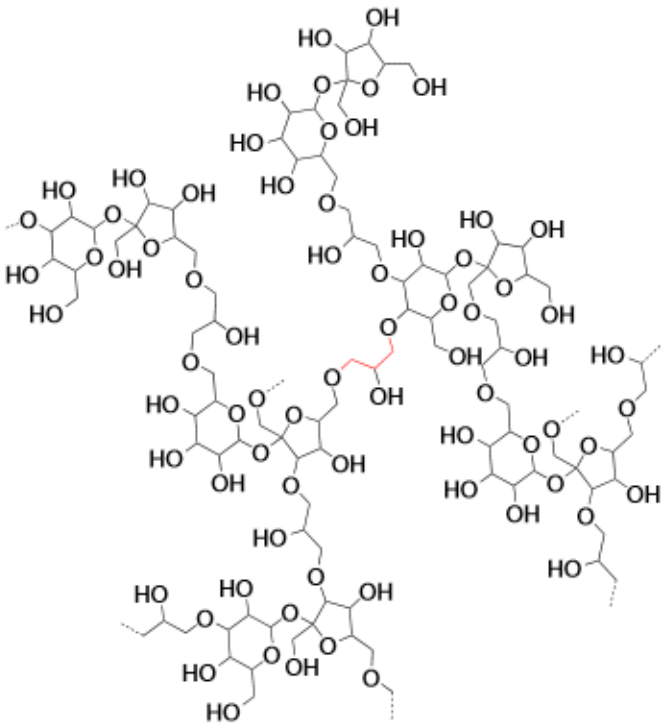


Figure 4.1 Ficoll is a highly cross-linked polysucrose epicholorohydin co-polymer. Representation of a Ficoll segment showing two chains cross-linked by epicholorohydrin. Ficoll has numerous hydroxyl groups which can participate in intermolecular hydrogen bonding.

For HX-MS analysis of ACTR in crowded conditions, we applied our recently developed strong cation exchange (SCX)-based cleanup to remove Ficoll before MS analysis.[15] ACTR is only transiently helical,[16, 17] so HX occurs rapidly at most ACTR backbone amides.[10, 18] To slow HX, we performed HX labeling at pH 5.5 allowing us to capture HX kinetics of ACTR amides over the time-course (seconds to hours) of a typical HX experiment. At 300 g L^{-1}, Ficoll caused destabilization of ACTR transient α-helices. In contrast, at 400 g L^{-1}, Ficoll caused stabilization of one ACTR transient α-helix. These results illustrate the complexity of interactions between proteins and polymer crowding agents and of crowding effects on IDPs in general.

4.2 Experimental

4.2.1 Protein expression and purification

ACTR, residues 1023-1093 (UniProt NCOA3_HUMAN), was co-expressed with residues 2059-2117 of the CREB binding protein (UniProt CBP_MOUSE) in *E. coli* [19] and purified using anion exchange and size-exclusion chromatography, as previously described.[18]. Assigned peptides are annotated as residues 1-71 where residue number one represents residue 1023. After purification, ACTR was concentrated by freeze-drying on a benchtop freeze drier (FreeZone, Labconco, Kansas City, MO) followed by reconstitution of the protein with ultrapure H_2O at one-third of the initial volume. The concentrated protein was then dialyzed into 20 mM sodium citrate, 50 mM sodium chloride, pH 5.5 buffer using 3 kDa cutoff MINI dialysis units (Thermo Scientific, West Palm Beach, FL). The final protein concentration was 46 μM ACTR by bicinchoninic acid assay using bovine serum albumin standards (Thermo Scientific, West Palm Beach, FL). The protein stock was divided into aliquots, flash frozen on liquid N_2, and stored at −80 °C. Crowded ACTR samples were prepared by adding solid Ficoll PM 70 (Sigma, St. Louis, MO) to thawed aliquots of ACTR stock, to reach 300 g L^{-1} or 400 g L^{-1} of Ficoll.

4.2.2 Hydrogen exchange labeling

HX labeling, quenching and injection was performed using a LEAP H/DX PAL robotic system with temperature-controlled compartments (LEAP Technologies, Carrboro, NC) configured for the SCX workflow as previously described.[15] HX labeling buffers were prepared with 20 mM sodium citrate, 50 mM sodium chloride in D_2O, pD 5.5 for ACTR labeling. Deuterated buffers are reported as pH adjusted to pD to correct for the deuterium isotope effect.[20] Crowded labeling buffers were the same citrate buffers with 300 g L^{-1} and 400 g L^{-1} of pre-deuterated Ficoll, prepared as described previously.[15]

To label ACTR in 300 g L^{-1} Ficoll, crowded ACTR samples (300 g L^{-1} Ficoll, pH 5.5) were labeled in a 19-fold excess of crowded labeling buffer (300 g L^{-1} pre-deuterated Ficoll, pD 5.5). For labeling in 400 g L^{-1} Ficoll, crowded ACTR samples (400 g L^{-1} Ficoll, pH 5.5) were labeled in a 9-fold excess of crowded buffer (400 g L^{-1} pre-deuterated Ficoll, pD 5.5). The corresponding uncrowded labeling experiments were performed similarly, in 19-fold and 9-fold uncrowded citrate buffers (pD 5.5), respectively. Undeuterated controls were labeled similarly in pH 5.5 citrated buffer using H_2O. After incubating for times between 13 s and 12 h, the labeled samples were quenched by diluting them 1:1 with 200 mM phosphate in H_2O (pH 2.5) at 1 ^oC and immediately injected into the LC stream for SCX cleanup and MS analysis.

4.2.3 Chromatography with SCX workflow

The automated SCX cleanup was performed in the refrigerated LC compartment of a H/DX PAL with valve switches and wash solution injections programmed in HDX Director software (LEAP Technologies, Carrboro, NC), as previously described.[15] Briefly, quenched HX samples were first digested using an immobilized pepsin column. The resulting peptides were trapped on an SCX trap while un-trapped Ficoll from crowded samples flowed to waste. The

SCX column was washed with chilled methanol/0.1 % formic acid to remove Ficoll contaminants. The SCX-trapped peptides were then eluted onto a C4 trap using chilled 0.5 M NaCl/0.1 % formic acid (pH 2.5) for desalting and concentration. After desalting, the peptides were eluted and separated on a C18 column using an acetonitrile/0.1% formic acid gradient flowing towards the MS for analysis. All SCX cleanup steps were completed in a few minutes at 1 $^{\circ}$C under quenched conditions.

4.2.4 Data Acquisition and Analysis

Peptic peptides were identified using accurate mass measurements (± 10 ppm) and MS/MS using collision-induced dissociation on a quadrupole time-of-flight mass spectrometer (Q-TOF, Agilent model 6530, Santa Clara, CA). Peptide MS and MS/MS spectra were assigned and analyzed using MassHunter Qualitative Analysis software (version B.7.0.0). A set of 20 peptides were found in common between crowded and uncrowded samples after labeling in both 300 g L^{-1} and 400 g L^{-1} Ficoll (see Table S1).

HX feature finding and peptide mass determination were performed using HD Examiner (Sierra Analytics, Modesto, CA). Deuterium uptake curves and standard deviations at each time-point for crowded and uncrowded labeling conditions were generated using an in-house R-script. The significance of differences between crowded and uncrowded uptake curves was determined by pooling standard deviations across all time points for each peptide in both conditions. The peptides were classified as significantly different if at least two HX labeling times were different after applying a two-sample Student's *t* test assuming equal variance at the 98% confidence interval, as described previously.[15]

The magnitude of statistically different HX rates between crowded and uncrowded conditions for each peptide was quantified using the following equation:

$$\Delta\overline{\mathrm{HX}} = \frac{\bar{m}_{\mathrm{c}} - \bar{m}_{\mathrm{u}}}{\bar{m}_{\infty} - m_0} \tag{4}$$

$\Delta\overline{\mathrm{HX}}$ represents the difference between the average deuterium uptake of a peptide across all HX times under crowded and uncrowded conditions, normalized by maximum observed deuterium uptake; the range for $\Delta\overline{HX}$ is between –1 and 1. $\bar{m}_{\mathrm{c}}$ and $\bar{m}_{\mathrm{u}}$ are the average masses over all HX labeling times for crowded and uncrowded conditions, respectively; $\bar{m}_{\infty}$ is the average mass after 12 h of HX, representing the totally deuterated control; and m_0 is the theoretical average mass of the undeuterated peptide.

4.3 Results

To determine the effects of crowding on ACTR, we labeled the protein in uncrowded D_2O buffer and in crowded D_2O buffers containing 300 and 400 g L^{-1} Ficoll. We recovered 20 ACTR peptides from samples containing 300 g L^{-1} and 400 g L^{-1} Ficoll, both representing 100 % coverage of the ACTR sequence (see Table 4.A and Figure 4.B). We have previously established that chemical exchange rates are unaffected by Ficoll in the labeling buffer,[15] so all differences in HX rates can be attributed to changes in ACTR structure under crowded conditions. For most peptides, deuterium uptake under crowded and uncrowded conditions converged by the 12 h labeling time indicating that there was equivalent HX in buffers with and without Ficoll (sees Figures 4.C and 4.D).

Figure 4.2 shows representative deuterium uptake plots for ACTR from crowded (blue, closed circles) and uncrowded (red, open circles) HX labeling. Figure 4.2A shows representative uptake plots comparing HX by ACTR in 300 g L^{-1} Ficoll with HX in uncrowded conditions and 4.2B shows uptake plots for the corresponding peptides comparing HX in 400 g L^{-1} Ficoll with HX in uncrowded conditions.

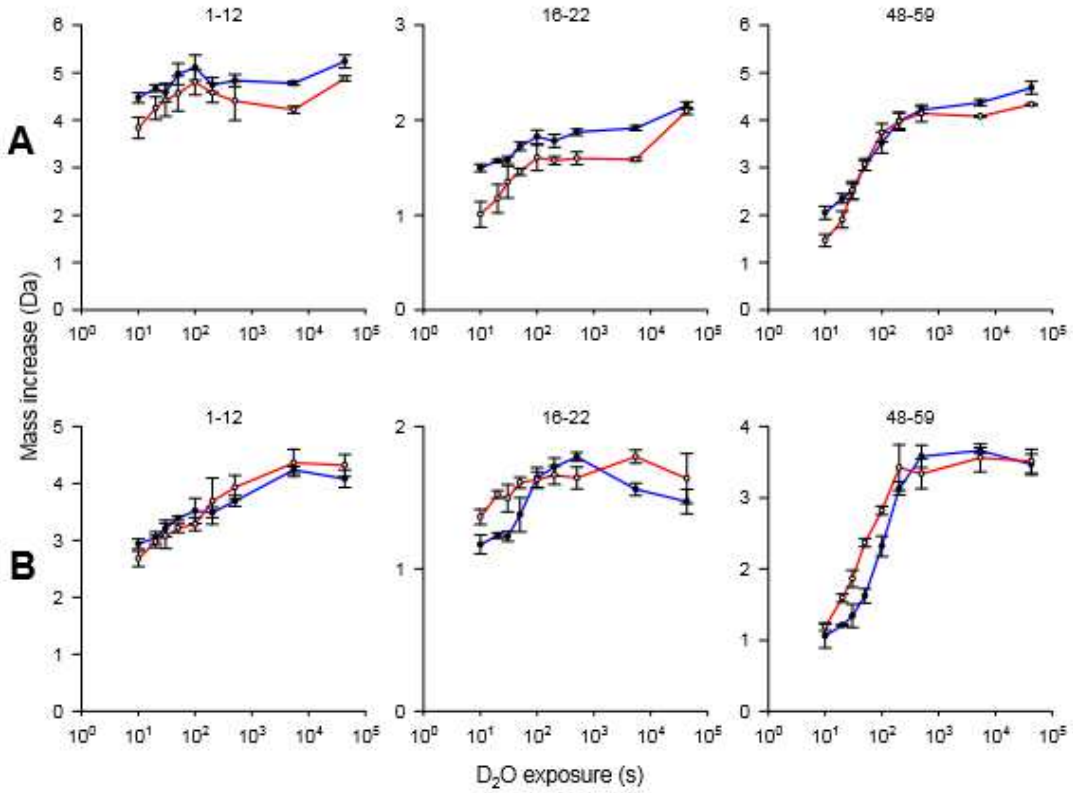


Figure 4.2 300 g L^{-1} Ficoll increases HX but 400 g L^{-1} Ficoll slows HX in some ACTR segments. Representative deuterium uptake plots for ACTR labeled in uncrowded (red, open circles) D_2O buffers of crowded (blue, closed circles) D_2O buffers. In row A, the crowded buffers contained 300 g L^{-1} Ficoll. In row B the crowded buffers contained 400 g L^{-1} Ficoll. Error bars represent one standard deviation for triplicate measurements; the numbers at the top of each plot denote peptide locations in the ACTR sequence. A complete set of all deuterium uptake curves is provided in Figures 4.C and 4.D of the Supporting Information.

We note here that the fraction of deuterium was different in 300 and 400 g L^{-1} Ficoll because the ACTR samples were diluted 1:19 (H_2O:D_2O) for 300 g L^{-1} measurements and 1:9 for 400 g L^{-1}. As a result, deuteration is higher in the peptides shown in Figure 4.2A than in the corresponding peptides in Figure 4.2B because the fraction of D is higher in the 300 g/L Ficoll samples.

300 g L^{-1} Ficoll caused two categories of differences in the HX kinetics of ACTR. Peptide 1-12 in Figure 4.2A is representative of ACTR regions where there were no significant differences in HX between crowded and uncrowded conditions. Peptides 16-22 and 48-59 in

Figure 4.2A are representative of ACTR regions where HX was faster in 300 g L^{-1} of Ficoll. The HX increase is greater in peptide 16-22 than in peptide 48-59. All uptake differences from the 20 peptides analyzed after labeling in 300 g L^{-1} Ficoll followed one of the trends represented by the plots in Figure 4.2A (see Figure 4.C). There are also two categories of HX differences between ACTR crowded by 400 g L^{-1} Ficoll and uncrowded conditions. Peptide 1-12 in Figure 4.2B is representative of ACTR regions where there were no significant differences in HX between crowded and uncrowded conditions. Peptide 16-22 in Figure 4.2B is representative of peptides where HX was not affected by 400 g L^{-1} Ficoll but where 300 g L^{-1} Ficoll caused faster HX. Peptide 48-59 in Figure 4.2B is representative of ACTR regions where 400 g L^{-1} Ficoll slowed HX but where 300 g L^{-1} Ficoll made HX faster. Figure 4.2B is representative of all differences in HX caused by 400 g L^{-1} Ficoll (see Figure 4.D).

To quantify the magnitude of the HX differences between crowded and uncrowded conditions, we calculated $\Delta\overline{\mathrm{HX}}$, as defined by equation (1). Here, $\Delta\overline{\mathrm{HX}} > 0$ indicates that Ficoll increased HX rates (i.e. deprotection from HX) and $\Delta\overline{\mathrm{HX}} < 0$ indicates that Ficoll decreased HX rates (i.e., protection from HX). The theoretical maximum and minimum for $\Delta\overline{\mathrm{HX}}$ are 1 and −1 respectively. Table 4.S2 lists $\Delta\overline{\mathrm{HX}}$ for all ACTR peptides that had statistically significant HX differences in 300 g L^{-1} and 400 g L^{-1} Ficoll relative to uncrowded conditions. Generally, 300 g L^{-1} Ficoll induced subtle deprotection ($0 < \Delta\overline{\mathrm{HX}} \leq 0.09$) of some ACTR regions while 400 g L^{-1} Ficoll induced only subtle protection ($0 > \Delta\overline{\mathrm{HX}} \geq -0.03$) of a smaller number of regions Figure 4.3 shows ACTR peptide segments with no significant differences (white), those that are destabilized (orange) and those that are stabilized (blue) by 300 g L^{-1} and 400 g L^{-1} Ficoll. Protection and deprotection from HX indicate stabilization and destabilization, respectively, of transiently helical regions of ACTR. In Figure 4.3, ACTR peptides are aligned with the ACTR

sequence and with a representation of regions that form transient α-helices (α1-3) in free ACTR based on secondary chemical shifts measured by NMR.[16, 17]

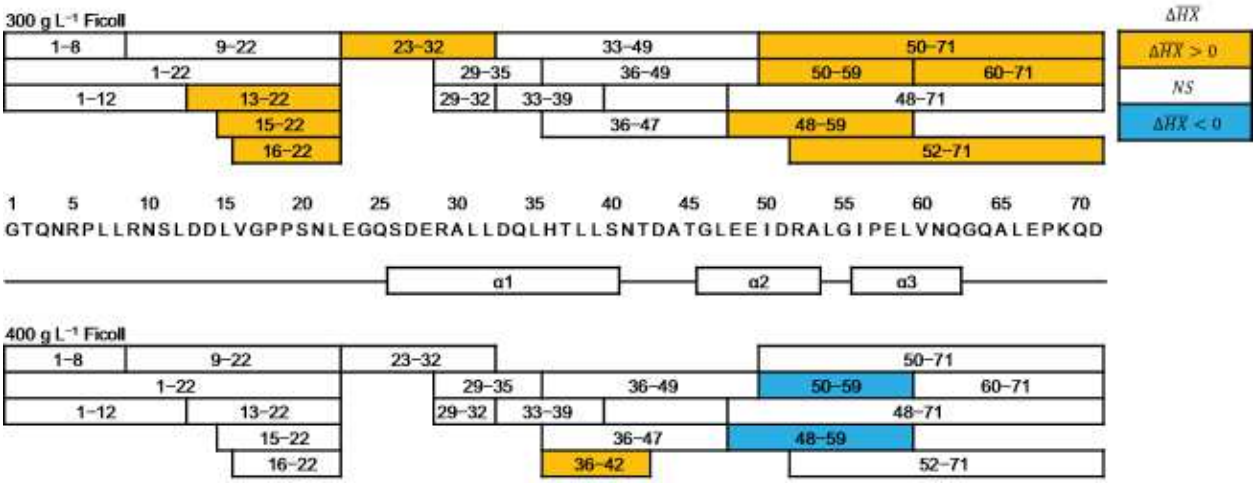


Figure 4.3 300 g L^{-1} Ficoll destabilizes but 400 g L^{-1} Ficoll stabilizes ACTR transiently helical regions. ACTR protection maps showing regions where Ficoll stabilizes (blue) and destabilizes (orange) ACTR. The transiently helical regions α1-3 are assigned based on ACTR structure in complex with CBP (PDB 1KBH).[19] The degree of protection (i.e. stabilization) or deprotection (i.e. destabilization) was determined by $\Delta\overline{HX}$ for peptides that showed statistically significant differences in uptake. $\Delta\overline{HX}$ was calculated according to equation (1). Residue 1 in the map corresponds to residue 1023 of the ACTR sequence (UniProt NCOA3_HUMAN).

In addition to the helical regions denoted in Figure 4.3, there is another segment, between residues 11-22, that is also weakly helical as predicted using the AGADIR helical propensity algorithm and as detected by NMR[16, 17] and millisecond HX.[10] Unstructured ACTR regions exchange quickly in the presence and absence of crowders (for example, 1-12 in Figure 4.2A) while transiently helical regions exchange more slowly in both conditions (see for example, 48-59 in Figure 4.2A). Generally, the HX differences between crowded and uncrowded conditions are less than the differences between unstructured and transiently structured regions indicating that crowding with Ficoll induces only subtle changes in ACTR transient helicity (compare Figures 4.C and 4.D). As shown in Figure 4.3, transiently helical regions of ACTR were destabilized by 300 g L^{-1} Ficoll, but only one transient helix was stabilized by 400 g L^{-1} Ficoll while the other regions were unaffected. Although it appears that peptide 36-42 was destabilized,

no other ACTR segments covering this region were destabilized by 400 g L^{-1} Ficoll and the HX difference is small ($\Delta\overline{HX} = 0.006$), thus we consider this result to be an outlier.

The N-terminal end of ACTR, spanning residues 1-12, was unaffected by 300 g L^{-1} Ficoll. Figure 4.3 shows that there was no difference in HX in all peptides that are associated with the N-terminal region (1-8, 1-12, 1-22 and 9-22). Peptides 1-8 and 1-12 were almost completely deuterated after only 13 s of labeling (see Figures 4.C and 4.D) suggesting that the N-terminal tail of ACTR is completely unstructured under all conditions. This observation is consistent with previously reported analysis of uncrowded ACTR [10, 16, 17] and also with our observation that Ficoll does not affect HX in unstructured protein segments of another IDP.[15] 300 g L^{-1} Ficoll destabilized the adjacent, weakly helical region between residues 11-22. Peptides from this region, 13-22, 15-22, and 16-22, were all destabilized in the presence of Ficoll. Two other peptides spanning this region, 1-22 and 9-22, had similar HX kinetics as peptide 13-22, but their HX differences were not statistically significant (see Figure 4.C and Table 4.E). This similarity in HX kinetics indicates that segments 1-22 and 9-22 cover the weakly helical region within segment 13-22, but because Ficoll-induced destabilization is relatively subtle, the expected increase in HX would be diluted by the additional N-terminal residues in peptides 1-22 and 9-22. We also note that region 13-22, which has the weakest helical propensity, experiences the greatest destabilization by Ficoll ($\Delta\overline{HX} = 0.09$ for peptide 16-22 in 300 g L^{-1} Ficoll). This result makes sense because we expect that weaker helices would be more prone to destabilization by Ficoll.

At 300 g L^{-1}, Ficoll destabilized the N-terminal edge of helical region 1 (residues 26-40). Figure 4.3 shows that peptide 23-32 is the only segment associated with α-1 that was destabilized in the presence of Ficoll. We observed similar HX kinetics for other peptides in this region (29-

32, 29-35, 33-39, 33-49, 36-47 and 36-49), but the differences were not statistically significant (see Figure 4.C and Table 4.E). The relatively slower HX kinetics in these peptides, compared to HX kinetics in peptides covering the flexible N-terminal, are consistent with transient helicity in the middle region of ACTR. The C-terminal helices, α-2 and α-3, spanned by residues 46-62, are more destabilized than α-1 by 300 g L^{-1} Ficoll. Most of the peptides in this region (48-59, 50-59, 50-71, 52-71 and 60-71) were destabilized in the presence of Ficoll. Peptide 48-71 covering α-2 and α-3 had similar HX kinetics as the other peptides in this region, but the HX difference was not statistically significant (see Figure 4.C and Table 4.E). ACTR region α-1 has the strongest helical propensity [10, 16, 17] explaining why α-1 is the least destabilized by Ficoll.

At 400 g L^{-1}, Ficoll caused stabilization of α-2, in contrast with the effect of Ficoll at 300 g L^{-1} where α-2 was destabilized. While in other regions, there were no significant differences in HX between crowded and uncrowded conditions (see Figure 4.3). Strikingly, 400 g L^{-1} Ficoll stabilized a small segment in the center of α-2: Figure 4.3 shows that peptides 48-59 and 48-71 were stabilized in the presence of Ficoll. Considering that the first two residues of each peptide lose their deuterium label during the LC step and by averaging out the residues from the overlapping peptides with no HX differences, we conclude that stabilization only occurs at residues 50-51. To further support this conclusion, we considered that peptide 60-71 had no HX difference in 400 g L^{-1} Ficoll, indicating that the stabilization in segment 48-71 is localized between residues 50-59. While 300 g L^{-1} Ficoll destabilizes ACTR transient secondary structure, surprisingly 400 g L^{-1} Ficoll appears to stabilize transient helicity in ACTR. We explain these unexpected results by considering hard and soft crowding interactions between ACTR and Ficoll which arise in highly concentrated Ficoll solutions.

4.4 Discussion

4.4.1 Macromolecular crowding effects

Many macromolecular crowding effects on proteins are adequately described in terms of hard (i.e., steric) interactions, also called volume exclusion. Volume exclusion theory states that the volume occupied by macromolecular crowding agents is inaccessible to the protein of interest.[3, 21] In solutions containing high concentrations of macromolecular crowding agents, proteins experience steric repulsions that limit their conformational flexibility. Volume exclusion theory has been employed (wholly or in part) to simulate macromolecular crowding with results that match experimental observations.[3, 4, 22, 23] The volume exclusion effect favors compact protein conformations and stabilizes natively structured proteins. An important prediction of the theory is that compaction of proteins increases with increasing concentrations of crowding agents. While the volume exclusion effect dominates macromolecular crowding of structured proteins, it is an incomplete picture of macromolecular crowding and is difficult to apply to IDPs that sample a diverse ensemble of conformations.[23] Experimental observations suggests that IDPs are more prone to soft crowding effects than to volume exclusion.[8]

Soft crowding effects occur because of non-specific chemical interactions between protein molecules and crowding agents. Soft interactions may be repulsive, attractive, or both, depending on the nature of the crowding agent and the protein of interest.[24, 25] Therefore, the balance between favorable and unfavorable soft interactions can either stabilize or destabilize compact protein conformations; the balance could also be concentration-dependent. An important distinction between soft interaction and hard interactions is that hard interactions are stabilizing and have a predictable concentration-dependence, at least within certain concentration limits. Another important difference is that soft interactions involve chemical contact (i.e.,

involve intermolecular forces) between the crowding agent and the protein. Chemical interactions like hydrogen bonding, dipolar interactions, electrostatic interactions, and hydrophobic interactions can influence protein secondary structure and tertiary structure by offering favorable *intermolecular* contacts between proteins and crowding agents to substitute for *intramolecular* contacts in the folded state.[26] Non-specific interactions can tune or even dominate volume exclusion effects especially for IDPs because IDPs have a preference for extended conformations.[8] Furthermore, the concentration-dependence of non-specific interactions with polymeric crowding agents is more complicated to predict because the arrangement of polymer chains is drastically different in different concentration regimes.[6] For macromolecular crowding studies of IDP it is particularly important to consider the concentration regime of polymeric crowding agents.

4.4.2 Polymer solutions for macromolecular crowding

The concentration-dependence of both volume exclusion and soft interactions on proteins results from the arrangements of polymer chains in dilute and semi-dilute concentration regimes.[6] In dilute solutions, polymer chains are isolated and move independently. Isolated polymers can be treated as inert, impenetrable objects that simply exclude volume. In the absence of soft interactions with the polymer, proteins in such a solution can only experience hard particle exclusion. As polymer concentration increases, however, the individual chains begin to experience mutual steric exclusion. At a critical limit known as the overlap concentration ($c*$), polymer chains cannot self-avoid by steric effects so that they make contacts. The overlap concentration of a specific polymer depends on its molecular weight (M_{w}) and the third power of its radius of gyration (R_{g}) according to equation (5)[27]

$$c^* = \frac{M_w}{\frac{4}{3}\pi R_g^3 N_A} \quad (5)$$

where N_A denotes, Avogadro's number. Beyond the overlap concentration ($c > c*$) the polymer solution is semi-dilute. In the case of linear polymers, the chains become intertwined in a mesh-like network.[6] The polymer network is characterized by an average distance between participating polymer chains, known as the mesh size (ξ). Mesh size determines the dynamic properties of the polymer chains themselves and of other macromolecules that may be trapped in the mesh, for example, the diffusion rate of a protein that is trapped within a cross-linked hydrogel.[28, 29] Moreover, the polymer mesh presents more opportunities for any proteins, embedded within the mesh, to make non-specific contacts with the crowding agent.[7]

Ficoll 70, is highly crosslinked and is better described as a semi-rigid sphere (or star polymer[30]) than as a linear polymer.[11, 12] Based on the molecular weight, approximately 200 sucrose monomers are contained in each Ficoll molecule; all within a radius of only 4.7 nm.[31] Based on these specifications, Ficoll molecules exist in solution as self-contained spheroidal polymer meshes whose interior resembles a semi-dilute solution of short linear polymers, as depicted in Figure 4.4. Xu et al proposed a similar picture for a highly cross-linked polymer.[32] However, Ficoll is not a rigid sphere; molecular sieve experiments suggest that Ficoll is much more flexible in solution than globular proteins of similar Stoke's radius.[33] In addition to being deformable, Ficoll comprises a complex distribution of conformations[12, 31] and molecular weights in solution. Owing to this flexibility and polydispersity, we expect Ficoll to arrange itself differently in 300 g L^{-1} and 400 g L^{-1} solutions.

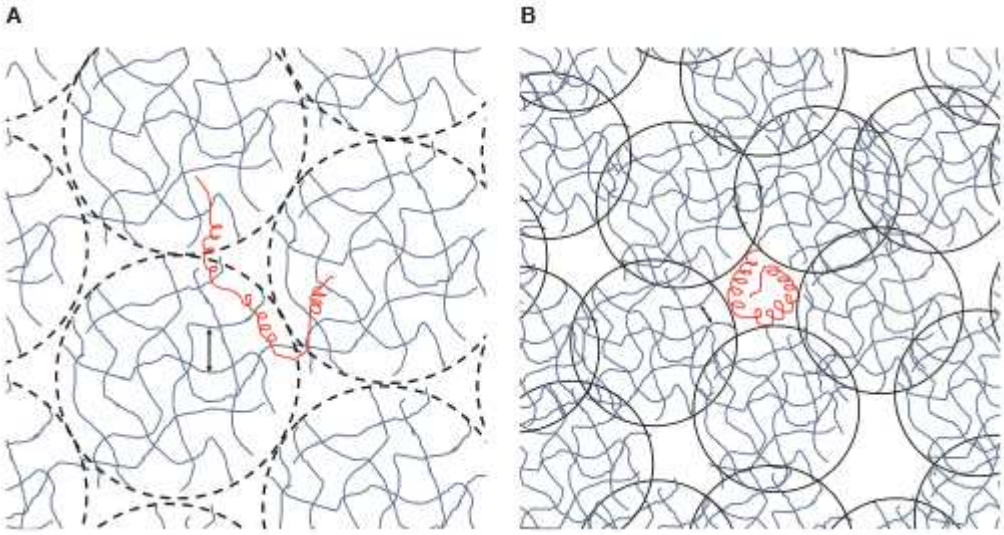


Figure 4.4 Ficoll changes structure in critical vs. semi-dilute concentration regimes inducing contrasting effects on ACTR transient helicity. Arrangement of Ficoll chains (blue) in relation to ACTR (red). Double arrows show the length (ξ) of the Ficoll mesh. (A) Expanded, penetrable Ficoll spheroids in the critical regime ($\boldsymbol{c} \approx \boldsymbol{c}*$) where ACTR penetrates the Ficoll mesh leading to destabilization of ACTR transient helices. (B) Compacted, impenetrable Ficoll spheroids in the semi-dilute regime ($\boldsymbol{c} > \boldsymbol{c}*$) where ACTR is squeezed out into the interstitial spaces. ACTR becomes compacted and transient helices are stabilized.

Using equation (5) and a reported hydrodynamic radius $R_g = 4.7$ nm [31], we estimate c^* for Ficoll to be 270 g L^{-1}. From that value, we surmise that 300 g L^{-1} represents a Ficoll concentration that is near the critical overlap concentration ($\boldsymbol{c} \approx \boldsymbol{c}*$) and that 400 g L^{-1} represents a semi-dilute solution ($\boldsymbol{c} > \boldsymbol{c}*$). Unlike linear polymers, and because Ficoll is deformable, we expect Ficoll to become both more compact and also more entangled in semi-dilute solutions, beyond $\boldsymbol{c}^*$. This compaction of Ficoll under the influence of volume exclusion is similar to, but greater in extent than, the effect of Ficoll on structured proteins[13] because Ficoll is less rigid than globular proteins. Volume exclusion will decrease both the radius of Ficoll and also the mesh size within each Ficoll molecule. The reduced mesh size would also affect the occupancy, and perhaps exclude other macromolecules from the Ficoll mesh (see Figure 4.4).

4.4.3 Polymer crowding effects on IDPs

In general, macromolecular crowding effects bear the mark of the crowding agent of choice, as much as they do the protein of interest. Monteith et al showed that protein quinary structure, arising from matrix-protein contacts can influence protein stability in cells.[26] *In vitro*, similar concentrations of globular protein crowding agents like albumin have different crowding effects than artificial polymer crowding agents like polyvinylpyrrolidone (PVP) on ubiquitin.[34] As we have described in the previous section, polymers also introduce nuanced crowding effects based on the identity of the polymer chosen as a crowding agent. The concentration regime of the polymer solution is one key determinant of crowding effects. Indeed, Soranno et al note that most polymer solutions used in crowding studies are in the semi-dilute regime.[8] Besides the polymer concentration, chain size and shape also factor into overall crowding effects. Moreover, natively structured proteins are affected differently than IDPs. Structured proteins, stabilized by intramolecular contacts, tend to avoid contact with polymer crowders even in semi-dilute polymer solutions and are always stabilized, as observed by Miklos et al using CI2 in different PVP solutions.[24] On the other hand, IDPs are likely to engage extensively in chemical contacts with polymer chains because IDPs, being unstructured, lack favorable intramolecular contacts and therefore they can freely engage with the polymer mesh.[8] The strength of the non-specific interactions depends on the chemical nature of the polymer. For example Ficoll can hydrogen bond far more extensively with proteins than polyethylene glycol (PEG) or PVP because Ficoll has many hydrogen bond donors and acceptors (see Figure 4.1).

Indeed, some unexpected effects of macromolecular crowding have been observed for different combinations of IDPs and crowders. Ficoll and dextrans did not induce any changes in the disordered domains of c-Fos and p27kip1 when probed by CD, intrinsic florescence, and 1-

anilinonaphthalene-8-sulfonate (ANS) binding, yet excluded volume theory predicts that compact conformations should be favored.[14] On the other hand, the radius of gyration of different IDPs, measured using FRET, decreased in crowded PEG solutions.[8] Volume exclusion theory was inadequate to quantitatively explain the observed extent of IDP compaction. These *in vitro* results along with in-cell measurements[35, 36] demonstrate that macromolecular crowding effects on IDPs are more nuanced than those on structured proteins. Herein, we have examined crowding effects of Ficoll on the structure and dynamics of the disordered protein ACTR, in two different Ficoll concentration regimes.

4.4.4 ACTR as a model for crowding effects on IDPs

ACTR is a random coil IDP with a propensity to form transient α-helices in three segments under dilute conditions (see Figure 4.3).[10, 16, 17, 19] ACTR is well-characterized in uncrowded solutions; here we are interested in how ACTR helical propensity is affected in crowded solutions containing Ficoll. We found that all transiently helical segments of ACTR were destabilized in 300 g L^{-1} Ficoll. In contrast, only one helical segment was stabilized in 400 g L^{-1} Ficoll. These unexpected results suggest that non-specific interactions are predominant near the overlap concentration ($c \approx c*$) of Ficoll but that volume exclusion becomes predominant in the semi-dilute regime ($c > c*$). Sorrano et al investigated conformational behavior of ACTR in crowded solutions containing PEG, another commonly-used polymer crowding agent.[8] To place our results in the context of crowding theory we begin with a brief summary of their findings.

Using single molecule FRET, Soranno et al found that the radius of gyration of ACTR (and other IDPs) decreased with increasing concentrations and with increasing PEG molecular weight.[8] The compaction behavior of ACTR with increasing PEG concentration matches

qualitative expectations, but the compaction trend with increasing PEG size does not match quantitative prediction of volume exclusion theory. The observed trends can be explained by the formation of a mesh network that can engage an IDP. In semi-dilute PEG solutions, non-specific interactions between ACTR and the PEG mesh modulate the compaction of ACTR. However, changes in ACTR transient helicity were not detectable by the single molecule FRET measurements of Sorrano et al. because single molecule FRET cannot directly probe secondary structure. To bridge this gap, we have used HX-MS to analyze ACTR structure with peptide resolution in different concentrations of Ficoll.

4.4.5 Contrasting effects on ACTR structure in different Ficoll concentration regimes

Ficoll-induced destabilization of ACTR transient helicity is expected near the overlap concentration of Ficoll because non-specific interactions between ACTR and Ficoll will dominate volume exclusion. In 300 g L^{-1} solutions ($c \approx c*$), Ficoll spheroids are expanded, but form an overlapping mesh. Because ACTR is a flexible, extended chain, segments of ACTR can infiltrate Ficoll and become *temporarily* incorporated in the Ficoll mesh. This is possible because Ficoll spheres have ample solvent-filled space. The density of a 4.7 nm Ficoll spheroid is approximately 0.3 g cm^{-3}, only about 20% of the density of crystalline sucrose,[37] suggesting that mutual volume exclusion between Ficoll molecules when $c \approx c*$ is insufficient to appreciably deform or compact the Ficoll spheres. Figure 4.4 illustrates how ACTR segments can engage with the mesh inside the Ficoll sphere. Unlike other polymers such as PEG and PVP, Ficoll contains numerous hydroxyl groups that act as both hydrogen bond donors and acceptors (see Figure 4.1), affording many opportunities for ACTR to form intermolecular hydrogen bonds with Ficoll hydroxyl groups.

The intermolecular hydrogen bonds do not affect the HX rates in the flexible N-terminal region of ACTR (or in unstructured peptides[15]). Soft interactions between ACTR and Ficoll manifest themselves as increased HX rates only in transiently helical regions of ACTR where there is competition between intramolecular (ACTR-ACTR) and intermolecular (ACTR-Ficoll) contacts. Amide protons that would otherwise be protected from HX by intramolecular hydrogen bonds in transiently helical regions in uncrowded solvent become more exchange competent in 300 g L^{-1} Ficoll because of soft interactions with Ficoll. Helix formation is less favorable in IDPs than in folded proteins and the mesh, unlike free solvent, presents a semi-rigid scaffold of hydroxyl groups so that the transient helices of ACTR molecule embedded in the Ficoll mesh are more destabilized by soft interactions with Ficoll. Furthermore, α-1, with the greatest helical propensity,[10, 16, 17] is more resistant to destabilization by Ficoll than the weaker helical regions 11-22, α-2 and α-3.

Our results may appear, at first, to contradict the observation by Sorrano et al who report that ACTR becomes more compact in highly concentrated solutions of PEG. However, ACTR interacts differently with Ficoll than it does with PEG because Ficoll is spherical and PEG is linear. As a result, Ficoll is arranged differently than PEG both near the critical concentration and in semi-dilute solutions. The available data suggest that Ficoll resembles globular protein crowding agents more closely than either extended or overlapping PEG chains. Also, Ficoll possesses hydroxyl functional groups whereas PEG is a polyether. Zhou et al note that PEG participates in favorable hydrophobic interactions with proteins leading them to recommend Ficoll as a preferred alternative for probing volume exclusion effects on proteins.[3] However, IDPs like ACTR lack hydrophobic side chains and so ACTR is unlikely to engage in favorable hydrophobic interactions with PEG. Finally, volume exclusion is always at play in the ACTR-

Ficoll system but the combination of favorable non-specific interactions between ACTR and Ficoll and the inherent flexibility of ACTR may overpower the tendency for ACTR to become compact in 300 g L^{-1} Ficoll in the same way it becomes compacted in 300 g L^{-1} of PEG.

The Ficoll-induced effects on transiently helical regions of ACTR in 400 g L^{-1} Ficoll are in stark contrast to those in 300 g L^{-1} Ficoll because volume exclusion effects dominate non-specific interaction in the higher Ficoll concentration. Not only are transiently helical regions of ACTR not destabilized in the semi-dilute ($c > c*$) Ficoll solution, but a segment in α-2 is actually stabilized in this concentration regime (see Figure 4.3). Again we consider the arrangement of Ficoll that is illustrated in Figure 4.4. Because Ficoll is deformable and because volume exclusion effects between Ficoll molecules increase in the semi-dilute regime, Ficoll spheroids can shrink in diameter.[32] The mesh size within Ficoll spheres would also shrink towards a limit where solvent-filled space between Ficoll chains becomes inaccessible to ACTR segments. At this limit, Ficoll spheres become impenetrable and ACTR is *squeezed-out* into the interstitial space between Ficoll spheres as shown in Figure 4.4B. This effect is similar to what happens in size-exclusion chromatography. The polymer matrixes that are used in size-exclusion chromatography are deliberately concentrated to exclude molecular species according to their ability to overcome steric (or volume exclusion) effects.[38]

We expect ACTR to become more compact in semi-dilute Ficoll because of volume exclusion effects within the interstitial spaces between the now impenetrable Ficoll spheres that ACTR is forced to occupy. Under these conditions, the effect of competing favorable soft interactions is drastically decreased so that the helical propensities of ACTR are no longer affected by soft interactions. The somewhat surprising result is that Ficoll begins to act like an inert crowder. The resulting compaction stabilizes a limited segment of ACTR because ACTR is

a random coil IDP with only weak helical propensity. Reduction in the radius of gyration of ACTR, in agreement with the observations of Soranno et al[8], does not necessitate localized stabilization of other transient helices because ACTR does not fold cooperatively. Moreover, we may be observing stabilization in only one region of ACTR in 400 g L^{-1} Ficoll because volume exclusion stabilizes or destabilizes protein secondary structures more subtly compared to non-specific interactions; and that level of subtlety may be below the limits of HX-MS detection. Previous investigations also revealed that crowding with Ficoll does not cause the expected stabilization of IDP transient structures.[14, 39]

4.4.6 300 g L^{-1} Ficoll also destabilizes α-helices of CBP, a molten globular IDP

We previously reported the crowding effects of Ficoll on CBP in 300 g L^{-1} Ficoll observed using HX-MS.[15] Unlike ACTR, CBP is a molten globule comprised of three stable α-helices but also a highly flexible tertiary structure.[16, 19, 40] As with ACTR, helical regions of CBP were destabilized by Ficoll in solutions that contained 300 g L^{-1} Ficoll. Again, CBP helices were destabilized because of non-specific interactions between CBP and Ficoll in the Ficoll mesh. Despite being molten globular, CBP is still highly flexible and can penetrate Ficoll spheres to engage in CBP-Ficoll hydrogen bonds that can compete with intramolecular contacts. Interestingly, one helical segment of CBP was also stabilized by 300 g L^{-1} Ficoll. This contrasting behavior between CBP and ACTR is expected because CBP is a molten globule which resembles globular proteins more than ACTR does. CBP can be more easily compacted than ACTR by volume exclusion in 300 g L^{-1} Ficoll because of stronger intramolecular contacts. All the same, destabilization of CBP helices by Ficoll was both more frequently observed (8

deprotected peptides vs 2 protected peptides) and larger in magnitude than stabilization ($-0.02 \leq \Delta\overline{HX} \leq 0.1$).

4.5 Summary and Implications

Our observations illustrate the complexity of predicting macromolecular crowding effects on IDPs. The native-like conformations of globular protein ensembles are predicted to become more stable in the presence of any crowding agent via volume exclusion. However, the unstructured IDP conformational ensembles are affected differently by different crowding agents because non-specific interactions can overcome both weak intramolecular interactions and volume exclusion effects. In other words, the flexibility of IDPs makes them more sensitive than structured proteins to their crowding environments because IDPs can engage more extensively with crowding agents. Cellular crowding agents can be globular, like structured proteins and protein complexes, or extended polymeric molecules like other IDPs, polypeptides, nucleotides, and polysaccharides. These crowding agents will also be distributed differently in different cellular compartments. Transient secondary or tertiary structures of IDPs may exhibit different stabilities depending on their local cellular environments. Therefore, the local cellular environment may affect the ability for an IDP to bind with protein ligands by altering the proportions of conformations in the IDP ensemble that are favorable for binding.

For *in vitro* crowding studies of IDPs the choice of polymer crowding agent is an important consideration that will depend on the crowding conditions to be simulated. Ficoll, for example, is spherical and resembles a globular crowding agent more than a linear polymer like PEG. However, Ficoll is more deformable than a globular protein and we expect the meshed interior of Ficoll spheres to be more accessible than the hydrophobic core of a globular crowding protein. In addition Ficoll does not present either charged or hydrophobic surfaces. Our

observations suggest that Ficoll acts like a highly-branched polymer in lower concentration regimes but like a hard sphere in semi-dilute regimes. This explanation is in line with the proposed approach of Soranno et al[8], that the polymeric natures of both the crowding agent and the IDP must be accounted for when making quantitative predictions of crowding effects. Indeed, the observed Ficoll-induced stabilization and destabilization of ACTR and CBP are more easily explained by considering the Ficoll polymer and its arrangement in different concentration regimes, as opposed to describing Ficoll exclusively as a hard sphere.

In conclusion, polymer systems for crowding studies of IDPs can be useful for simulating cellular crowding only after careful consideration of the chemical and physical properties of the polymer and the IDP. The functional groups of the monomer units that make up the polymer will determine the type and strength of non-specific interactions. A polyether like PEG is will engage in hydrophobic interactions while a polysaccharide like Ficoll or dextran will engage in hydrogen-bonding interactions. Since IDP sequences are depleted in hydrophobic residues,[41, 42] IDPs can form more favorable interactions with crowding agents like Ficoll than with crowders like PEG. Our observations provide insights into the behavior of IDPs at the secondary structure level in crowded environments. An interesting next step would be to perform HX-MS analyses of ACTR in solutions containing different crowding agents. Results of such analyses would broaden our understanding of the stability of ACTR transient helices inside cells, a more complex crowded environment than a synthetic polymer solution. In general, the findings we have presented here help to narrow the gap in our understanding of how the interplay between simple volume exclusion effects and complex non-specific interaction affects IDP secondary structure in crowded environments.

4.6 References

[1] Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci. 11*, 739-756.

[2] Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli, *J. Mol. Biol. 222*, 599-620.

[3] Zhou, H.-X., Rivas, G., and Minton, A. P. (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences, *Annu. Rev. Biophys. 37*, 375-397.

[4] Elcock, A. H. (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment, *Curr. Opin. Struct. Biol. 20*, 196-206.

[5] Zhou, H.-X., and Dill, K. A. (2001) Stabilization of proteins in confined spaces, *Biochemistry 40*, 11289-11293.

[6] de Gennes, P.-G. (1979) *Scaling concepts in polymer physics*, Cornell University Press, Ithaca, NY.

[7] Koutsioubas, A., Lairez, D., Combet, S., Fadda, G., Longeville, S., and Zalczer, G. (2012) Crowding effect on helix-coil transition: beyond entropic stabilization, *J. Phys. Chem. 136*, 215101-215108.

[8] Soranno, A., Koenig, I., Borgia, M. B., Hofmann, H., Zosel, F., Nettels, D., and Schuler, B. (2014) Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments, *Proc. Natl. Acad. Sci. U.S.A. 111*, 4874-4879.

[9] Hvidt, A., and Nielsen, S. O. (1966) Hydrogen exchange in proteins, *Adv. Protein Chem. 21*, 287-386.

[10] Keppel, T. R., and Weis, D. D. (2015) Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging, *J. Am. Soc. Mass. Spectrom. 26*, 547-554.

[11] Venturoli, D., and Rippe, B. (2005) Ficoll and dextran vs. globular proteins as probes for testing glomerular permselectivity: effects of molecular size, shape, charge, and deformability, *Am. J. Physiol.- Renal 288*, F605-F613.

[12] Fissell, W. H., Manley, S., Dubnisheva, A., Glass, J., Magistrelli, J., Eldridge, A. N., Fleischman, A. J., Zydney, A. L., and Roy, S. (2007) Ficoll is not a rigid sphere, *Am. J. Physiol.- Renal 293*, F1209-F1213.

[13] Dhar, A., Samiotakis, A., Ebbinghaus, S., Nienhaus, L., Homouz, D., Gruebele, M., and Cheung, M. S. (2010) Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding, *Proc. Natl. Acad. Sci. U.S.A. 107*, 17586-17591.

[14] Flaugh, S. L., and Lumb, K. J. (2001) Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27Kip1, *Biomacromolecules 2*, 538-540.

[15] Rusinga, F. I., and Weis, D. D. (2016) Automated strong cation exchange cleanup to remove macromolecular crowding agents for protein hydrogen exchange mass spectrometry, *Anal. Chem. 89*, 1275-1282.

[16] Ebert, M.-O., Bae, S.-H., Dyson, H. J., and Wright, P. E. (2008) NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR†, *Biochemistry 47*, 1299-1308.

[17] Kjaergaard, M., Nørholm, A. B., Hendus–Altenburger, R., Pedersen, S. F., Poulsen, F. M., and Kragelund, B. B. (2010) Temperature□dependent structural changes in intrinsically

disordered proteins: Formation of α–helices or loss of polyproline II?, *Protein Science 19*, 1555-1564.

[18] Keppel, T. R., Howard, B. A., and Weis, D. D. (2011) Mapping unstructured regions and synergistic folding in intrinsically disordered proteins with amide H/D exchange mass spectrometry, *Biochemistry 50*, 8722-8732.

[19] Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators, *Nature 415*, 549-553.

[20] Glasoe, P. K., and Long, F. A. (1960) Use of glass electrodes to measure acidities in deuterium oxide1,2, *J. Phys. Chem. 64*, 188-190.

[21] Minton, A. P. (1981) Excluded volume as a determinant of macromolecular structure and reactivity, *Biopolymers 20*, 2093-2120.

[22] Hall, D., and Minton, A. P. (2003) Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges, *Biochim. Biophys. Acta 1649*, 127-139.

[23] Minton, A. P. (2005) Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: macromolecular crowding and protein stability revisited, *Biophys. J. 88*, 971-985.

[24] Miklos, A. C., Li, C., Sharaf, N. G., and Pielak, G. J. (2010) Volume exclusion and soft interaction effects on protein stability under crowded conditions, *Biochemistry 49*, 6984-6991.

[25] Sarkar, M., Lu, J., and Pielak, G. J. (2014) Protein crowder charge and protein stability, *Biochemistry 53*, 1601-1606.

[26] Monteith, W. B., Cohen, R. D., Smith, A. E., Guzman-Cisneros, E., and Pielak, G. J. (2015) Quinary structure modulates protein stability in cells, *Proc. Natl. Acad. Sci. U.S.A. 112*, 1739-1742.

[27] Rubinstein, M., and Colby, R. H. (2003) *Polymer Physics*, Oxford University Press, New York, NY.

[28] Buenger, D., Topuz, F., and Groll, J. (2012) Hydrogels in sensing applications, *Progress in Polymer Science 37*, 1678-1719.

[29] Pescosolido, L., Feruglio, L., Farra, R., Fiorentino, S., Colombo, I., Coviello, T., Matricardi, P., Hennink, W. E., Vermonden, T., and Grassi, M. (2012) Mesh size distribution determination of interpenetrating polymer network hydrogels, *Soft Matter 8*, 7708-7715.

[30] Huber, K., Bantle, S., Burchard, W., and Fetters, L. J. (1986) Semidilute solutions of star branched polystyrene: a light and neutron scattering study, *Macromolecules 19*, 1404-1411.

[31] Georgalis, Y., Philipp, M., Aleksandrova, R., and Krüger, J. (2012) Light scattering studies on Ficoll PM70 solutions reveal two distinct diffusive modes, *J. Colloid Interface Sci. 386*, 141-147.

[32] Xu, F., Kiba, Y., and Baba, Y. (2005) Nucleic Acids, Oligonucleotides, and DNA: Capillary Electrophoresis, In *Encyclopedia of Chromatography* (Cazes, J., Ed) Vol 2, pp 1120-1130, Taylor & Francis, Boca Raton, FL.

[33] Asgeirsson, D., Venturoli, D., Fries, E., Rippe, B., and Rippe, C. (2007) Glomerular sieving of three neutral polysaccharides, polyethylene oxide and bikunin in rat. Effects of molecular size and conformation, *Acta Physiologica 191*, 237-246.

[34] Wang, Y., Sarkar, M., Smith, A. E., Krois, A. S., and Pielak, G. J. (2012) Macromolecular crowding and protein stability, *J Am Chem Soc 134*, 16614-16618.

[35] Dedmon, M. M., Patel, C. N., Young, G. B., and Pielak, G. J. (2002) FlgM gains structure in living cells, *Proc. Natl. Acad. Sci. U.S.A. 99*, 12681-12684.

[36] McNulty, B. C., Young, G. B., and Pielak, G. J. (2006) Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder, *J. Mol. Biol. 355*, 893-897.

[37] Haynes, W. M., Ed. (2011) *CRC Handbook of Chemistry and Physics*, 91st ed., Taylor Francis Group.

[38] Hagel, L. (2001) Gel-Filtration Chromatography, In *Current Protocols in Molecular Biology* (Ausubel, F.M., Ed.) 44, pp 10.9.1-10.9.2, Greene Publishing Associates, Brooklyn, NY.

[39] Mukherjee, S., Waegele, M. M., Chowdhury, P., Guo, L., and Gai, F. (2009) Effect of macromolecular crowding on protein folding dynamics at the secondary structure level, *J. Mol. Biol. 393*, 227-236.

[40] Kjaergaard, M., Teilum, K., and Poulsen, F. M. (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP, *Proc. Natl. Acad. Sci. U.S.A. 107*, 12535-12540.

[41] Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting protein disorder for N-, C-and internal regions, *Genome Inf. 10*, 30-40.

[42] Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins: Struct., Funct., Bioinf. 42*, 38-48.

Appendices

4.A ACTR peptide assignments

Table 4.A ACTR peptide assignments in common between crowded and uncrowded conditions

Peptide*	Sequence	Retention time	Observed m/z	Charge	Mass (observed)	Mass (calculated)	Mass error (ppm)
1-8	GTQNRPLL	7.476	449.7592	2	897.5032	897.5032	-0.06
1-12	GTQNRPLLRNSL	7.487	456.9299	2	1367.7667	1367.7633	2.42
1-22	GTQNRPLLRNSLDDLVGPPSNL	8.508	792.7606	3	2375.2593	2375.2557	1.52
9-22	RNSLDDLVGPPSNL	8.268	748.8902	2	1495.769	1495.7631	3.96
13-22	DDLVGPPSNL	7.86	513.7619	2	1025.5087	1025.5029	5.63
15-22	LVGPPSNL	7.587	398.7296	2	795.4464	795.4491	-3.32
16-22	VGPPSNL	7.213	683.373	2	682.3689	682.365	5.68
23-32	EGQSDERALL	7.615	559.2802	2	1116.5454	1116.5411	3.85
29-32	RALL	7.298	472.3264	1	471.319	471.3169	4.33
29-35	RALLDQL	8.176	414.7526	2	827.4903	827.4865	4.56
33-39	DQLHTLL	8.106	420.2361	2	838.4574	838.4549	2.99
33-49	DQLHTLLSNTDATGLEE	8.376	928.9491	2	1855.8887	1855.88	4.73
36-42	HTLLSNT	7.069	393.2123	2	784.4101	784.4079	2.76
36-47	HTLLSNTDATGL	7.833	621.8206	2	1241.6265	1241.6252	1.05
36-49	HTLLSNTDATGLEE	7.686	750.8628	2	1499.712	1499.7104	1.09
48-59	EEIDRALGIPEL	8.793	677.8671	2	1353.7193	1353.714	3.89
48-71	EEIDRALGIPELVNQGQALEPKQD	8.506	888.1293	3	2661.366	2661.361	1.89
50-59	IDRALGIPEL	8.561	548.8221	3	1095.6291	1095.6288	0.22
50-71	IDRALGIPELVNQGQALEPKQD	8.337	802.1014	2	2403.2811	2403.2758	2.21
52-71	RALGIPELVNQGQALEPKQD	8.182	726.0635	3	2175.1685	2175.1648	1.71
60-71	VNQGQALEPKQD	6.616	663.8361	2	1325.6576	1325.6575	0.06

*Residue 1 corresponds to residue 1023 of the ACTR (UniProt NCOA3_HUMAN, residues 1023-1093).

4.B ACTR peptic peptide map

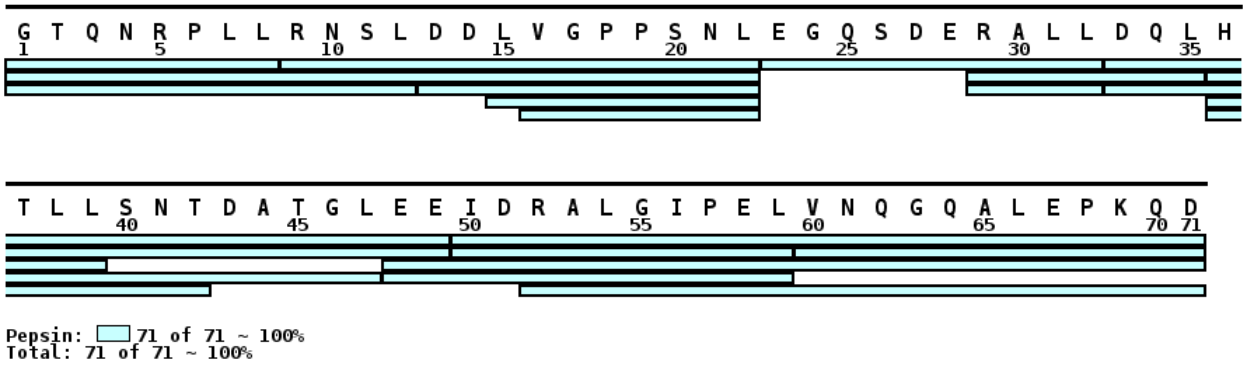


Figure 4.B ACTR peptic peptide map from crowded samples, containing 300 g L^{-1} Ficoll. 21 peptides, covering 100% of the ACTR sequence were recovered and used for HX-MS analysis, except peptide 36-42. In the 400 g L^{-1} Ficoll measurements, peptide 33-49 was not used. The map was generated using MS Tools [Kavan, D. and Man, P. "MSTools - Web based application for visualization and presentation of HXMS data" *Int. J. Mass Spectrom.* 2011, *302*, 53-58. http://dx.doi.org/10.1016/j.ijms.2010.07.030.]

4.C ACTR peptide deuterium uptake curves in 300 g L^{-1} Ficoll

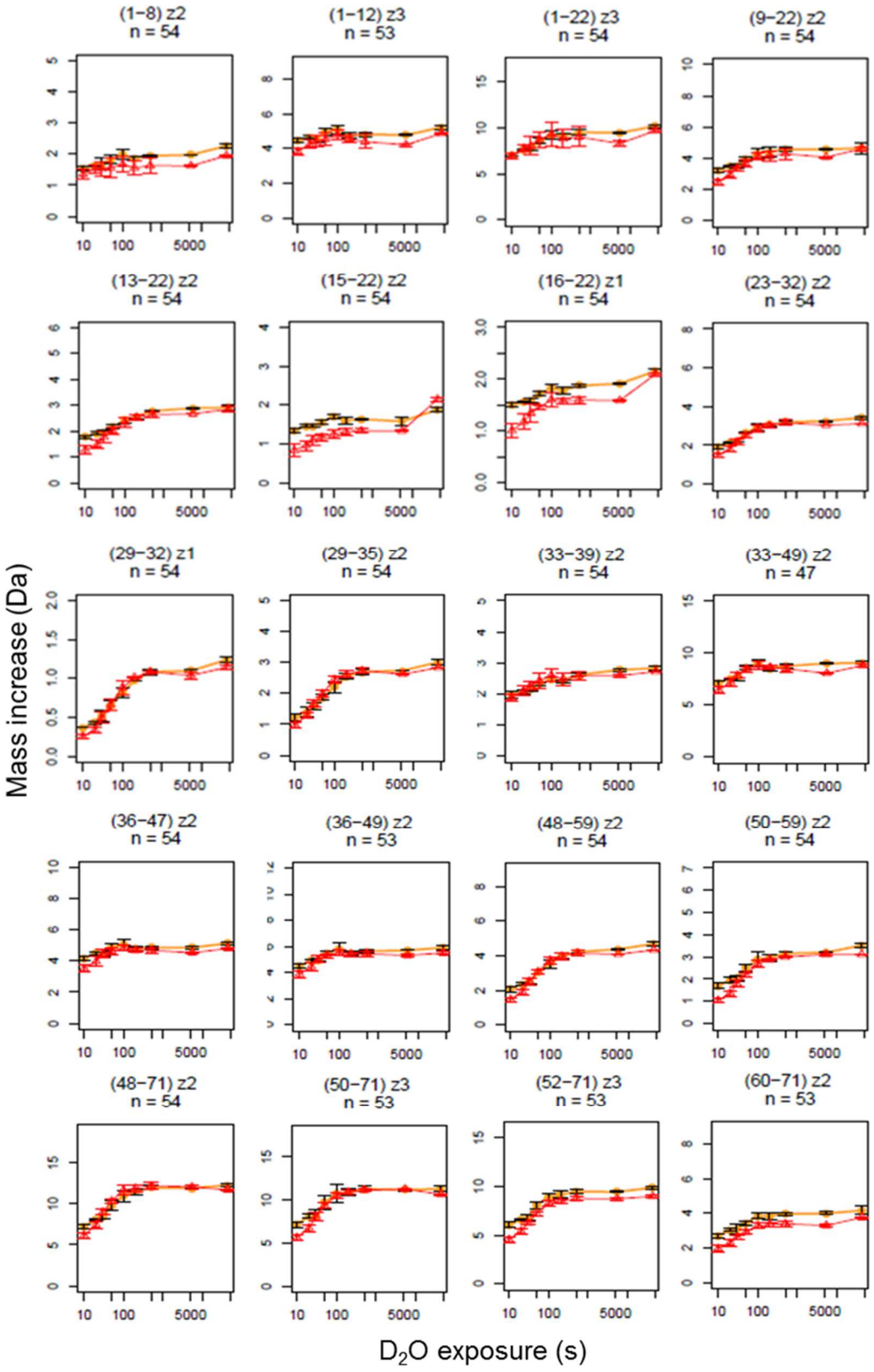


Figure 4.C ACTR uptake plots from 300 g L^{-1} Ficoll labeling showing uptake in red (uncrowded) and orange (crowded) buffers. The notation z indicates the peptide charge state and n denotes the total number of individual HX measurements depicted on each plot from a total possible of 54.

4.D ACTR peptide deuterium uptake curves in 400 g L^{-1} Ficoll

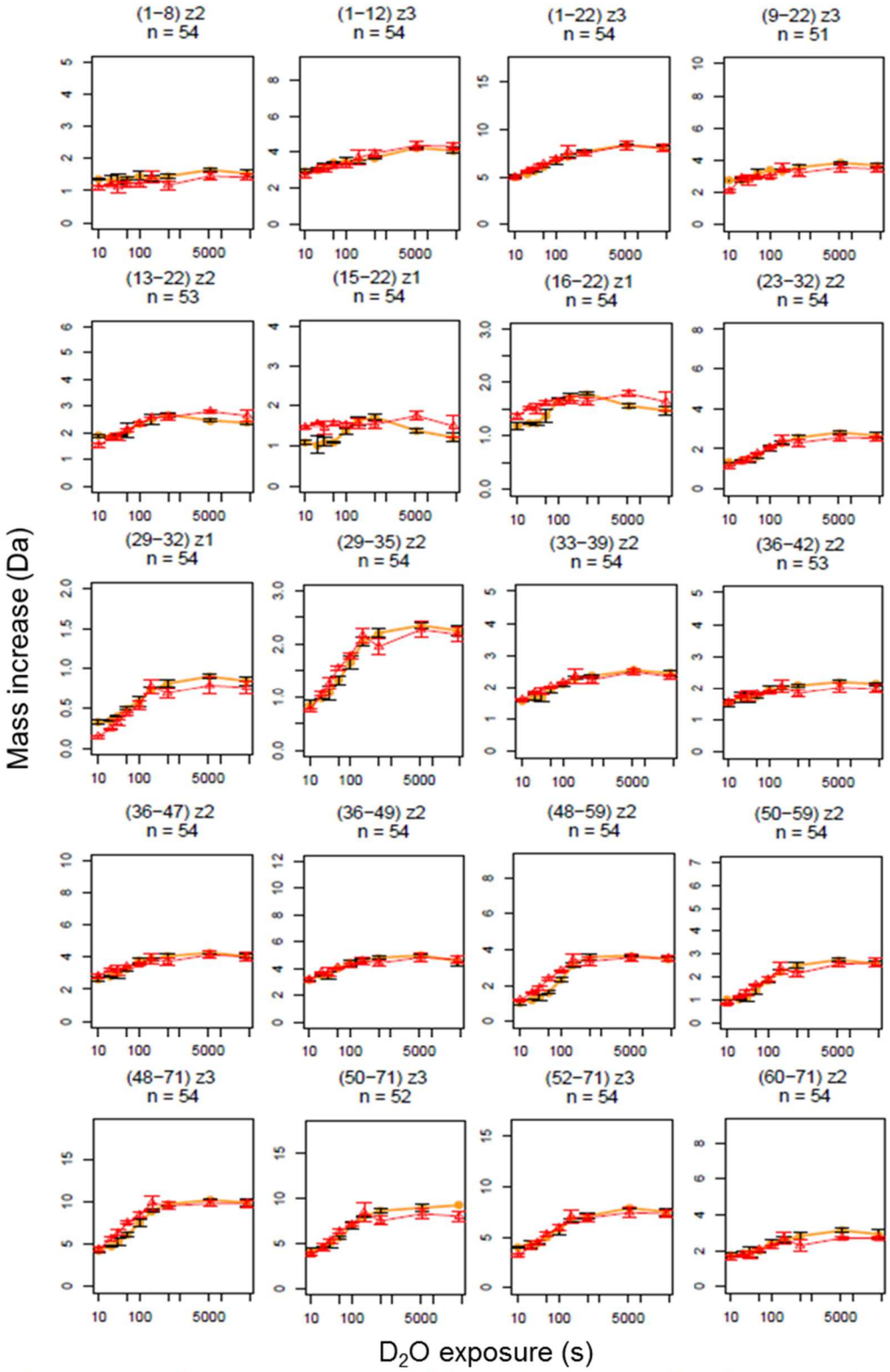


Figure 4.D ACTR uptake plots from 400 g L^{-1} Ficoll labeling showing uptake *in red* (uncrowded) and orange (crowded) buffers. The notation z indicates the peptide charge state and n denotes the total number of individual HX measurements depicted on each plot from a total possible of 54.

4.E $\Delta\overline{HX}$ of ACTR peptides in the two crowding conditions

Table 4.E $\Delta\overline{HX}$* of ACTR peptides in the two crowding conditions

Peptide	$\Delta\overline{HX}$ in 300 g/L Ficoll	$\Delta\overline{HX}$ in 400 g/L Ficoll
1-8	NS	NS
1-12	NS	NS
1-22	NS	NS
9-22	NS	NS
13-22	0.03	NS
15-22	0.08	NS
16-22	0.09	NS
23-32	0.02	NS
29-32	NS	0.02
29-35	NS	NS
33-39	NS	NS
33-49	NS	Not detected
36-42	Not detected	NS
36-47	NS	NS
36-49	NS	NS
48-59	0.02	−0.04
50-59	0.04	NS
48-71	NS	−0.04
50-71	0.03	NS
52-71	0.05	NS
60-71	0.06	NS

*$\Delta\overline{HX}$ was calculated using equation (1). NS denotes no significant difference based on criteria described in the Data Acquisition and Analysis subsection of the Experimental Procedures.

Chapter 5: Conclusions, Updates & Future Directions

5.1 Structural basis for activation of calcineurin by calmodulin

5.1.1 Revised mechanism for calcineurin activation

Calcineurin (CaN) is a heterodimeric phosphatase with a catalytic CaN A subunit and a regulatory CaN B subunit. In the presence of high calcium levels, CaN is activated by binding of CaN B and calmodulin (CaM), which are both calcium-dependent, to their respective binding site on CaN A.[1-3] The CaM binding site is located in an unstructured region of CaN A, called the regulatory domain (RD).[4] The C-terminal end of the RD is connected to an autoinhibitory domain (AID), which is helical when it is located in the active site cleft of CaN A.[5] CaN A terminates with a disordered C-terminal tail (CT). Chapter 2 describes HX-MS experiments on a CaN-RD-AID-CT construct in the absence and in the presence of CaM. HX-MS confirms that free CaN-RD-AID-CT is largely unstructured. Upon binding CaM, the RD folds in the CaM binding domain and in an immediately-adjacent region on the C-terminal end of the CaM binding domain. CD data suggests that the folding in the RD involves the formation of α-helices. HX in the AID and CT were unaffected by CaM binding. From the HX-MS data, a mechanism for calcineurin activation was proposed involving interactions between CaM and two regions of the RD, the CaM-binding domain and the immediately-adjacent region, called the *distal helix*.[6] Interactions with CaM, stabilize α-helices in those two regions, and displaces the AID from the active site cleft of CaN A. The HX data also suggests that the AID is unstructured (outside the active site cleft). Displacement of the AID exposes the active site, activating CaN A.

Following the study outlined in chapter 2, HX-MS experiments were performed on free and CaM-bound full-length CaN heterodimer (both CaN A and CaN B subunits) in the presence of calcium.[7] HX-MS data on the heterodimer confirmed that the RD-AID-CT region of CaN A is

also unstructured in the heterodimer. Furthermore, the data confirmed the CaM-induced folding in the CaM-binding domain and in the distal helix. In addition, CaM-induced structural changes were also observed in other CaN domains. Surprisingly, the AID remained unchanged in the free and CaM-bound CaN heterodimer, which appears to contradict the X-ray crystal structure showing a helical AID in the CaN A active site.[5] This result, regarding the AID, matches the HX-MS results with the RD-AID-CT construct. The HX changes seen in other regions of CaN, led to a revision of the CaM-dependent mechanism for calcineurin activation. In addition to the formation of an α-helix in the CaM-binding domain, the distal helix and the C-terminal end of the CaN B subunit interact with CaM, anchoring it to the CaM binding site. The CaM binding interactions result is slightly increased flexibility in the AID, which could result from its displacement from the active site cleft. This updated model for CaN activation shows a possible limitation of the reductionist approach, i.e. using only the RD-AID-CT construct, described in chapter 2. Moreover, HX results on CaN heterodimer reveal that the multistep activation of CaN is complex and requires further investigation in different conditions, for example, in the absence of calcium.

5.1.2 Macromolecular crowding effects on calcineurin activity

More recently, calcineurin was studied in solutions containing the synthetic crowding agents dextran and Ficoll.[6] The results showed that crowding stabilizes the α-helix in the region of the RD, adjacent to the CaM-binding domain on the C-terminal end. This is the same region, observed in HX-MS experiments to undergo structural changes upon CaM binding. This (so called) *distal helix* is also involved in the proposed mechanisms for CaN activation. Furthermore, the stabilization of distal helix increases calcineurin activity. These results confirm the importance of the distal helix in CaN activation and suggest another level of calcineurin

regulation, involving macromolecular crowding. The study of crowding effects on the unstructured CaN RD is an example of macromolecular crowding studies on IDPs, the topic of discussion in chapters 3 and 4 of this dissertation.

5.2 Macromolecular crowding effects on IDP transient structure

5.2.1 Summary of crowding effects on CBP and ACTR

CBP is a molten globular IDP with stable α-helices but a highly flexible tertiary structure. Ficoll, a synthetic crowding agent, stabilized some CBP regions but destabilized the edges of CBP α-helices. Ficoll-induced stabilization was expected because molten globules resemble ordered proteins (more than random coil IDPs) and should therefore experience compaction due to volume exclusion. On the other hand, destabilization of helical segments was also expected because of non-specific interactions between CBP and Ficoll chains. Specifically, CBP and Ficoll can engage in intermolecular hydrogen-bonding, which can substitute intramolecular hydrogen-bonding, especially at less stable helix edges. The disruption of helical segments is even more pronounced for ACTR in a Ficoll solution similar to the one used in CBP experiments. ACTR is a random coil IDP with transiently helical regions, i.e. weaker helices than CBP. In 300 g L^{-1} Ficoll, all transiently helical regions of ACTR were destabilized due to non-specific interactions. Interestingly, 400 g L^{-1} Ficoll did not destabilize ACTR transient helicity. Instead, one transiently helical regions of ACTR was stabilized. This result is attributed to changes in the arrangement of Ficoll molecules, at higher concentrations, results in ACTR compaction due to volume exclusion. Moreover, this results agrees with some observations from studies of ACTR in crowded solutions containing PEG[8] but it also expands on those observations by showing the localizing crowding effects on IDP structures in different IDP-crowder pairs.

5.2.2 Future direction: IDPs and synthetic crowders

Contrasting observations have been made between effects of different crowding agents on proteins.[9] For example, protein crowders, such as albumin, showed different extents of volume exclusion effects on CI2, a globular protein, than synthetic crowders, such as Ficoll.[10] These differences occur because of the complex interplay of volume exclusion and non-specific interactions between proteins and different crowding agents. For globular proteins, volume exclusion is the predominant crowding effect, although non-specific interactions also occur, which can modulate volume exclusion effect. On the other hand, IDPs tend to have extended conformations, affording them more opportunities to engage in non-specific interactions with crowders. IDPs were observed to have unexpected concentration-dependent compactions behaviors, which could not be accounted for by volume exclusion theory alone.[8] Those observations, together with HX-MS results discussed in chapter 4, suggest that the IDP-crowder pair is a key determinant of overall crowding effect. Given the heterogeneity of the intracellular crowded environment, it could be useful to perform crowding experiments using different polymer crowders then comparing potentially contrasting effects on a particular IDP, such as ACTR. The SCX-based HX-MS method, described in chapter 3, would facilitate relatively quick data acquisition from crowding experiments using a variety of crowding agents. Based on comparisons of HX rates and taking into account the different physical and chemical properties of different crowders, inferences could be drawn about crowding effects on IDP structure in different cellular compartments. Another approach, which could also provide useful information, would be to use mixtures of different polymer crowders, which could simulate heterogeneous crowding conditions in a single experiments. Results from such experiments could be compared with those from non-crowded conditions or crowded conditions with individual crowders.

Furthermore, datasets from many different IDP-crowder combinations could provide useful parameters for developing effective computer models for crowding effects on IDPs.

5.2.3 IDPs and biological macromolecular crowders

While solutions containing polymer crowders can be convenient for simulating cellular crowding, the extent to which polymers can compared to physiologically relevant crowders is debatable. Ideally, protein or biomolecular crowders would be used in all crowding studies. A simple setup has been reported for removing nucleic acids from protein samples under quench conditions, using a strong anion exchange column to extract negatively charges nucleic acid from intact proteins.[11] It would also be interesting to develop equally simple approaches to remove protein crowders for HX-MS analysis of IDPs. With respect to the SCX-based workflow for HX-MS, as outlined in chapter 3, an obvious disadvantage would be the overwhelming amount of peptides from a protein crowder, perhaps a thousand-fold more concentrated than the protein of interest. These peptides are likely to outcompete peptides of interest for binding to the SCX column and desalting trap. Crowder peptides would also co-elute and suppress ionization of peptides of interest, making HX-MS analysis impractical. In spite of the impracticality of the SCX setup (per se), for HX-MS with protein crowders, the same basic approach could be implemented using different columns, combinations of columns, automated timing of valve switches and pump gradients, and wash procedures. A possible experiment could involve using proteolysis-resistant proteins, such as lysozyme, to crowd IDPs. Since IDPs are relatively more prone to proteolysis than lysozyme, IDP peptides could then be separated from the intact protein crowder. That separation could be based on differences of retention on SCX or other columns, such as reversed-phase. In short, a variety of different, commercially available or in-house HPLC columns, can be employed to exploit physico-chemical differences between globular proteins

and IDPs for experimental approaches similar to the SCX-based method. Moreover, similar principles can be extended to crowding studies using other biological molecules, such as nucleic acids and polyssacharides, as crowding agents.

5.3 References

[1] Stemmer, P. M., and Klee, C. B. (1994) Dual calcium ion regulation of calcineurin by calmodulin and calcineurin B, *Biochemistry 33*, 6859-6866.

[2] Yang, S.-A., and Klee, C. B. (2000) Low affinity Ca2+-binding sites of calcineurin B mediate conformational changes in calcineurin A, *Biochemistry 39*, 16147-16154.

[3] Aramburu, J., Rao, A., and Klee, C. B. (2001) Calcineurin: from structure to function, *Curr. Topics Cell. Reg. 36*, 237-295.

[4] Shen, X., Li, H., Ou, Y., Tao, W., Dong, A., Kong, J., Ji, C., and Yu, S. (2008) The secondary structure of calcineurin regulatory region and conformational change induced by calcium/calmodulin binding, *J. Biol. Chem. 283*, 11407-11413.

[5] Kissinger, C. R., Parge, H. E., Knighton, D. R., and Lewis, C. T. (1995) Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex, *Nature 378*, 641.

[6] Cook, E. C., and Creamer, T. P. (2016) Calcineurin in a crowded world, *Biochemistry*.

[7] Al-Naqshabandi, M. A. H. (2016) Conformational analysis of intrinsically disordered proteins using mass spectrometry-based approaches, University of Kansas.

[8] Soranno, A., Koenig, I., Borgia, M. B., Hofmann, H., Zosel, F., Nettels, D., and Schuler, B. (2014) Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments, *Proc. Natl. Acad. Sci. U.S.A. 111*, 4874-4879.

[9] Sarkar, M., Li, C., and Pielak, G. J. (2013) Soft interactions and crowding, *Biophys. Rev. 5*, 187-194.

[10] Benton, L. A., Smith, A. E., Young, G. B., and Pielak, G. J. (2012) Unexpected effects of macromolecular crowding on protein stability, *Biochemistry 51*, 9773-9775.

[11] Sperry, J. B., Wilcox, J. M., and Gross, M. L. (2008) Strong anion exchange for studying protein–DNA interactions by H/D exchange mass spectrometry, *J. Am. Soc. Mass. Spectrom. 19*, 887-890.